

ICES WKADSAM REPORT 2010

SCICOM STEERING GROUP ON SUSTAINABLE USE OF ECOSYSTEMS

ICES CM 2010/SSGSUE:10

REF. SCICOM, ACOM

Report of the Workshop on Reviews of Recent Advances in Stock Assessment Models World- wide: "Around the World in AD Models" (WKADSAM)

27 September - 1 October 2010

Nantes, France



ICES
CIEM

International Council for
the Exploration of the Sea

Conseil International pour
l'Exploration de la Mer

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

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

Recommended format for purposes of citation:

ICES. 2010. Report of the Workshop on Reviews of Recent Advances in Stock Assessment Models Worldwide: "Around the World in AD Models" (WKADSAM), 27 September - 1 October 2010, Nantes, France. ICES CM 2010/SSGSUE:10. 122 pp.
<https://doi.org/10.17895/ices.pub.9053>

For permission to reproduce material from this publication, please apply to the General Secretary.

The document is a report of an Expert Group under the auspices of the International Council for the Exploration of the Sea and does not necessarily represent the views of the Council.

© 2010 International Council for the Exploration of the Sea

Contents

Executive summary	1
1 Introduction	3
1.1 Terms of Reference (ToRs).....	3
1.2 Intended approach.....	3
1.3 Terminology	4
1.4 Report structure	4
2 Software packages and themes	5
2.1 SAM.....	5
2.1.1 Description.....	5
2.1.2 Summary of WKADSAM discussion.....	7
2.2 BREM (Two-stage Biomass Random Effects Model)	8
2.2.1 Description: Application of BREM to Bay of Biscay anchovy	8
2.3 Stock Synthesis 3 (version 3.11b)	13
2.3.1 Description.....	13
2.3.2 Summary of WKADSAM discussion	24
2.4 MULTIFAN-CL.....	25
2.4.1 Description.....	25
2.4.2 Summary of WKADSAM discussion	28
2.5 CASAL	29
2.5.1 Description.....	29
2.5.2 Summary of WKADSAM discussion	29
2.6 TINSS.....	30
2.6.1 Description.....	30
2.6.2 Summary of WKADSAM discussion	31
2.7 CSA	32
2.7.1 Description.....	32
2.7.2 Summary of WKADSAM discussion	33
2.8 ADAPT-VPA	34
2.8.1 Description.....	34
2.8.2 Summary of WKADSAM discussion	37
2.9 ASAP	37
2.9.1 Description.....	37
2.9.2 Summary of WKADSAM discussion	44
2.10 SURBA.....	45
2.10.1 The development of SURBA	45
2.10.2 The SURBA method.....	46
2.10.3 Summary of WKADSAM discussion	50
2.11 XSA 51	
2.11.1 Description.....	51

2.11.2	Summary of WKADSAM discussion.....	51
2.12	B-ADAPT	51
2.12.1	Description.....	51
2.12.2	Summary of WKADSAM discussion.....	52
2.13	General issues in benchmarks.....	52
2.13.1	Summary of presentation	52
2.13.2	Summary of WKADSAM discussion.....	61
2.14	Generic model features	62
2.14.1	Summary of presentation	62
2.14.2	Summary of WKADSAM discussion.....	65
3	The selection of modelling approaches and software packages.....	68
3.1	Experience with software packages	68
3.2	Case studies of model change: northern and southern hake.....	69
4	Conclusions	73
4.1	General recommendations	73
4.2	Recommendations for the 2012 Conference.....	75
5	References	76
	Annex 1: List of participants.....	79
	Annex 2: Software package descriptions	84

Executive summary

The Workshop on the Reviews of Recent Advances in Stock Assessment Models Worldwide “Around the World in AD Models” (WKADSAM) was convened in Nantes, France during autumn 2010 as the first meeting of the three-year ICES Strategic Initiative on Stock Assessment Methods (SISAM). Despite prevailing economic difficulties which have affected travel budgets for many institutes, the Workshop was successful in attracting participants from all over the world: 21 of the 32 attendees came from outside Europe.

The main interest for the Workshop lay in having a group of practicing stock assessment scientists and model developers discuss the models that are currently used around the world. The goal during the Workshop was to compare and contrast different modelling approaches in a systematic manner to provide guidance to ICES working groups on when particular models or approaches would be or would not be useful. The Workshop was specifically not a competition. In preparation for the meeting, a catalogue of models was prepared through the ICES Working Group on Methods of Stock Assessment (WGMG) as a starting point for discussions and included as an annex to this report. Because, the catalogue alone is not enough for colleagues to understand fully the practical details of a particular model, the presentation of case studies focussing on which problem each method has fixed was particularly important for helping to guide stock assessment scientists to a limited number of possible models to consider. Any conclusions were not to be prescriptive, but rather to present information clearly and fairly to allow informed model selection.

The first three days of the Workshop were taken up with presentations and discussions of methods and approaches. Much of the remaining two days was occupied with discussions about the hands-on experience of WKADSAM participants of the model packages presented, the important issue of model selection, the development of agreed terminology, and the generation of general conclusions from the meeting, along with recommendations for the forthcoming 2012 Conference. The general conclusions in brief are:

- WKADSAM recognizes the importance of the distinction that ICES has made between benchmark and update stock assessments. During the benchmark process for a given stock, a number of candidate *research models* should be applied to demonstrate the robustness of the advisory model. The *advisory model* (used for update assessments) should not be reviewed as part of the update advice process: the update assessments should only be subject to a technical audit. The purpose of the advisory model is not to understand every underlying real-world process but to provide robust advice.
- The order of importance for stocks for consideration in benchmark assessments should be: 1) stocks that are currently assessed incorrectly; 2) stocks that are not currently assessed and for which an assessment is required; 3) stocks for which the assessment could potentially be improved.
- The development of new stock assessment approaches should focus on situations where standard models cannot be applied, due to data or process constraints.
- New members of ICES assessment WGs (who have assessment responsibility) are encouraged to be able to write a simple stock assessment model as a minimum.

- WKADSAM does not recommend the use of one standard model package for ICES assessments, nor should all assessments use different methods. Selection of both the modelling approach and software package to be used for each stock should be based on a thoughtful consideration of the available data, biology of the stock, management requirements, statistical principles, and (importantly) available expertise.

1 Introduction

1.1 Terms of Reference (ToRs)

The ICES Workshop on Reviews of Recent Advances in Stock Assessment Models Worldwide “Around the World in AD Models” (WKADSAM) chaired by Coby Needle, UK*, and Chris Legault*, USA will meet in Nantes, France, 27 September to 1 October 2010 to collate, review and comment on stock assessment methods currently in use around the world.

This will be part of the ICES initiative on stock assessments methods.

The workshop was to:

- a) Determine the key techniques and approaches used to assess fish stocks
- b) Consider inter alia utility, ease of use, estimation procedures, robustness, suitability to different data richness, applicability to data poor situations, and relevance of assumptions in the models
- c) Summarize the advantages and disadvantages of the various methods, and describe the appropriate use.
- d) Comment on demonstrations by model developers of the utility of methods with case studies and simulated datasets, focussing on the question: What problem has the method fixed?
- e) Prepare the groundwork for a following workshop in 2011 or 2012 (see initiative plan below)

WKADSAM formed the first phase be part of the ICES Strategic Initiative on Stock Assessment Methods (SISAM). WKADSAM reported by January 2011 for the attention of the SISAM Steering Group.

1.2 Intended approach

In preparation for the meeting, it became clear that the ToRs were causing a degree of confusion among potential participants as to what exactly was to be achieved by the Workshop. To address this, the Chairs circulated a note which was intended to clarify the issues. This had implications for the outcomes of the meeting, and it is germane to summarize this note here, as follows.

This workshop was not to be a “Methods Olympics” where competing models battle for designation as the “best” model. There simply would not have been sufficient time during a week-long meeting to apply a large number of models to the large number of datasets that would be needed to ensure “fairness” (in a competitive sense). Such a competition would result in models that correspond most closely to the data and processes performing best, but the ability to generalize the results to all ICES assessments would be quite limited. The e-mail discussions preceding the Workshop served to demonstrate how difficult it is to set up such comparative tests appropriately.

Instead, the main interest lay in having a group of practicing stock assessment scientists and model developers discuss the models that are currently used around the world. A necessary first step for this discussion was to catalogue the currently available models, which the ICES Working Group on Methods of Fish Stock Assessment (WGMG) is currently undertaking. The goal during the workshop was to use this catalogue as a starting point for discussions. The catalogue alone is not enough for

colleagues to understand fully the practical details of a particular model, and uncovering these was always likely to fill a considerable part of the available time.

Although ToR c) may have misled people to think this meeting would be a model competition, the Chairs viewed it instead as a way to help ICES assessment working groups sort through the many available models outside those currently used for their particular assessment to find potential alternatives. The presentation of case studies showing what problem each method has fixed (ToR d) was particularly important for helping to guide stock assessment scientists to a limited number of possible models to consider. Any conclusions were not to be prescriptive, but rather present information clearly and fairly to allow informed model selection.

1.3 Terminology

Towards the end of the meeting, concerns were expressed by several participants that terms such as “model” and “software package” were being used interchangeably, and in a way that could confuse subsequent readers of the report. To address this, two definitions were made that will be used throughout:

- **Modelling approach:** set of equations and data that allow estimation of population abundance or other metrics of interest
 - e.g. statistical catch-at-age, virtual population analysis, time-series analysis
- **Software package:** a specific case of a modelling approach that is named
 - e.g. Stock Synthesis, CASAL, Multifan-CL, SAM, etc.

There was also considerable discussion over classifications of software-package families. WKADSAM settled on the following genre set:

- **Flexible, multipurpose:** packages which are intended to be applicable to a broad range of stock and data situations. Examples from the WKADSAM meeting include SS3, CASAL and Multifan-CL.
- **Specific, data issue-driven:** packages which can be applied to a number of stocks, but which are specific in their data requirements. Examples from the WKADSAM meeting include XSA, Adapt VPA, and many of the packages currently used by stock assessment WGs.
- **Custom, stock specific:** bespoke code which is written for a particular purpose and is not intended to be widely used.

The advantages and disadvantages of each of these families are discussed further throughout this report, and in particular in the Conclusions (Section 4).

1.4 Report structure

The bulk of this report is taken up by considerations of the software packages presented at the WKADSAM meeting (Section 2), which is intended to cover most of ToRs a) – d). Each subsection of Section 2 contains the extended summary of the model package presented (prepared by the presenter), along with a summary of the discussion following each presentation as collated by the relevant rapporteur. Several subsections also include a summary of the presentation itself from the rapporteur. **It should be noted that the material provided by the presenter of each model or package has been included in the report without any editing for content – therefore, some of the conclusions reached in these sections may not necessarily represent a consensus view.**

Section 3 summarizes the hands-on experience of WKADSAM participants of the model packages presented, and concludes that such experience is actually quite limited. It also covers an important case study for the issue of model selection, looking into the decisions taken during the recent benchmarking process for ICES hake stocks. Section 4 then offers general conclusions from the meeting, along with recommendations for the forthcoming 2012 Conference.

Finally, the available software package descriptions that have been collated by the ICES Working Group on Methods of Fish Stock Assessment (WGMG) during 2010 are brought together in a series of annexes. Not all of these packages were discussed during the WKADSAM meeting, but it is worthwhile to bring all the descriptions together in one place for future reference.

2 Software packages and themes

The following table summarizes the software package (or theme) presentations given at WKADSAM. For each of the packages, a description template was also filled in: these can be found in Annex 2, along with description templates for a number of packages in use in ICES and elsewhere that were not presented at WKADSAM.

SECTION	SOFTWARE PACKAGE / THEME	PRESENTER
2.1	SAM	Anders Nielsen
2.2	BREM	Verena Trenkel
2.3	Stock Synthesis 3	Rick Methot
2.4	MULTIFAN-CL	Shelton Harley
2.5	CASAL	Matt Dunn
2.6	TINSS	Steve Mattell
2.7	CSA	Benoit Mesnil
2.8	Adapt VPA	Chris Legault
2.9	ASAP	Chris Legault
2.10	SURBA	Coby Needle
2.11	XSA	Chris Darby
2.12	B-ADAPT	Chris Darby
2.12	General issues in benchmarks	Lionel Pawlowski
2.13	Generic model features	Doug Butterworth

2.1 SAM

2.1.1 Description

The state-space fish stock assessment model (SAM) was summarized to the group with focus on the rationale behind using random effects to describe the underlying random variables that are not observed (fishing mortalities and stock sizes). SAM is an age structured time-series model designed to be an alternative to the (semi-) deterministic procedures (e.g. VPA, Adapt, and XSA) and the fully parametric statistical catch at-age models (e.g. SCAA, and SMS). Compared to the deterministic procedures it solves the problem of falsely assuming catches-at-age are known without errors, and in addition the problem of selecting appropriate so-called 'shrinkage' and in certain cases convergence problems in the final years. Compared to fully parametric statistical catch at-age models SAM avoids the problem of fishing mortality being restricted to a parametric structure (e.g. multiplicative), and many problems related

to having too many model parameters compared to the number of observations (e.g. borderline identification problems, convergence issues, and asymptotic results).

In addition the model has a number of appealing properties. It allows selectivity to gradually evolve during the data period, it allows missing data, and finally it estimates the underlying process noise, which is useful for forward predictions.

Previous implementations of state-space assessment models (Gudmundsson 1987, 1994, and Fryer 2001) have been based on the extended Kalman filter, which uses a first order Taylor approximation of the non-linear parts of the model. The current implementation is based on the Laplace approximation which is better suited to handle non-linearities, and further validated by importance sampling.

It was presented how the recent decision to change XSA-shrinkage from 0.5 to 0.75 for Eastern Baltic Cod radically changed the perception of the stock in the final year to be more in line with the state space assessment model. The presenter argued against using ad-hoc criteria for setting these shrinkage parameters.

The state-space model has previously been validated at the methods working group by comparing to existing assessments and via simulated data. To further validate the model, it was extended to allow jumps in the underlying process to follow a mixture between a Gaussian and a fat-tailed Cauchy distribution, as opposed to a purely Gaussian. The model applied to North Sea Cod estimated the Cauchy fraction to be zero, and even forcing the Cauchy fraction to be 30% did not make the underlying process take noticeable sharper jumps. Finally a recent extension to allow the fishing mortality processes to be correlated was presented.

SAM is currently run for the following stocks in ICES: Kattegat Cod, Western Baltic Cod, Sole in 3A, Eastern Baltic Cod, North Sea Sole, Plaice in 3A, and North Sea Cod. Of these the state-space assessment model is primary for the first three stocks, and included as exploratory for the remaining. In addition to the stocks mentioned above it has been applied to other stocks (Western Baltic spring-spawning herring, North Sea Haddock, 3PS Cod, and Georges Bank Yellowtail Flounder) for testing purposes, and has performed well.

A simple web interface (<http://www.stockassessment.org>) to the state-space assessment model was presented. Collaboration at assessment working groups are often reduced to one or two members doing the actual assessment modelling, and remaining working group members reviewing and commentating on the results only. Part of the reason most working group members don't even try to reproduce the assessment, is that it takes a lot of work to get everything set up correctly. Typically several programs (specific versions) need to interact and the data need to be on a specific format. The web interface presented completely removes this obstacle. Once the stock coordinator has set up an assessment all members can reproduce the assessment and all the resulting graphs and tables simply by logging in and pressing 'run'. The working group members can also experiment with the model configuration and input data and easily compare the results. It would clearly be beneficial to have more hands and eyes on the details of each assessment.

Rapporteur's summary of presentation

SAM is an age-structured state-space model, with stochastic recruitment coming in every year (assuming some stock-recruitment relationship and lognormally distributed recruitment deviations). Log-normal process errors are assumed along cohorts. Fishing mortality follows a random walk in log-space with normally distributed in-

crements, applying, in principle, an independent random walk for each age. The data consist of catch numbers-at-age and survey indices-at-age, with observation error considered both in the catch and the survey indices. Model parameters consist of the variances of the process and observation errors, including the variance of the random walk for the $\log(\text{fishing mortality-at-age})$, the surveys' catchabilities and the parameters intervening in the stock-recruitment relationship. SAM is implemented in AD Model Builder, using its random effects module. A web interface has been developed for SAM, in order to facilitate its use. The model is used for several ICES stocks, either as the main assessment or as an alternative exploratory assessment. In order to allow for some potentially big jumps in F values in some time periods, the normal distribution for the $\log(F)$ increments can be replaced by a mixture of a normal and a t -distribution with low degrees of freedom. Another alternative explored for modeling $\log(F)$ is to have the normal increments correlated, instead of independent, across ages. Perfect correlation between the ages would correspond to separable fishing mortality, but with the annual factor of the fishing mortality following a random walk (in log-space) rather than being treated as a separate parameter for each year. Development of SAM continues and new features will be added.

2.1.2 Summary of WKADSAM discussion

The presentation was well received and the model was found to be useful and promising.

A remark was made about the difficulties (or near impossibility) often encountered when trying to estimate both process and observation error in state-space models. It is often found that the estimate of one of the variances goes to zero and, for example, fixing the ratio of both variances has sometimes been used as a "fix". The author said that the problem had not been encountered in the applications that have used SAM to date. A SAM like method was experimented with for sandeel, a species with short cohorts and very noisy data, and for that it was impossible to split observation and process error.

The question of whether a comparison had been performed between SAM and models where F is treated as a parameter, even if assigned a random walk distribution, was asked. The answer was negative. The importance of estimating the model parameters instead of using arbitrary values was highlighted.

Pros and cons of running the program directly on a web server were discussed. In particular, some people felt that this could be inconvenient as it required having Internet access. The author explained that all relevant files could be downloaded to one's own PC and run locally. He felt that having the web setting made things more clear and transparent and that the web interface could make things easier for non-experienced users.

The improved ability of the model to detect jumps in F by using a mixture of a normal and a t -distribution with low degrees of freedom was discussed. Some questions were raised about the ability of the model to detect such changes if these were of smaller magnitude than the ones in the example considered.

The author highlighted the fact that model fitting follows a clearly defined statistical procedure (maximum likelihood), hence avoiding difficult decisions like choosing an appropriate level of shrinkage in XSA.

Presently SAM does not allow inputting catches at the fleet level, only at stock level. It allows tuning fleets, but they are currently treated as surveys.

Work is going on at present to develop a multistock configuration of SAM.

SAM is a purely age-structured model. No length-structured configuration has been developed.

Robustness to poor ageing has not been examined and ageing errors are not explicitly modelled in SAM, but observation noise on age-classified catches is part of the model.

2.2 BREM (Two-stage Biomass Random Effects Model)

2.2.1 Description: Application of BREM to Bay of Biscay anchovy

The text provided by the presenter for this Section outlines the results of a case study. Details on the method itself can be found in the relevant table in Annex 2.

2.2.1.1 Data and model

Two time-series of biomass estimates for age 1 and total biomass were available for anchovy in the Bay of Biscay. The first one is obtained using the daily egg production method, referred to as DEPM series, and the second on using acoustics and identification trawl hauls, the acoustic series.

To make the model identifiable, one of the catchability parameters has to be fixed. Two choices are explored: 1) $q_{b^{acoustic}} = 1$; 2) $q_{b^{DEPM}} = 1$. Note that the recruit and total survey indices per method are assumed to have the same CV.

2.2.1.2 Results

Preliminary analyses

To check the consistency of the survey indices, a cohort plot of acoustic biomass indices-at-age was produced (Figure 1 left). The plot shows that successive indices of cohorts decreased as expected with the exception of the 2005 cohort for which the biomass index for age 1 was substantially smaller than for age 2 one year later (highlighted by a circle in figure 1). Indeed, survey indices for both the acoustics and the DEPM were unusually low for age 1 and also total biomass in 2005 (Figure 1 right). Thus it seems possible that the 2005 survey indices, at least for age 1, do not reflect stock biomass in the same manner as in other years. Given that the model assumes constant catchabilities across years, this would violate the assumptions of the model. Therefore an additional model fit was carried out where all data for 2005 was removed.

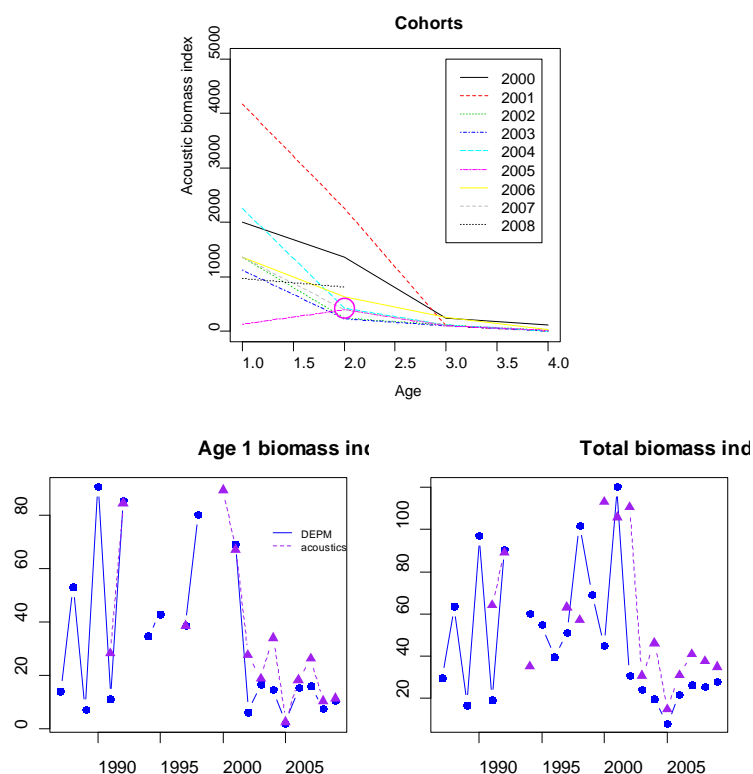


Figure 1. Cohort plot for acoustic biomass indices (left) and time-series of all survey indices (right).

Brem assumes that recruitment follows a lognormal distribution with no correlation between subsequent years. To check the validity of this assumption, the autocorrelation in survey indices for age 1 was calculated for the years 2000–2009 (Figure 2). Due to missing values it could not be calculated for earlier years. The results indicate that there was no autocorrelation in neither the acoustics nor the DEPM survey series.

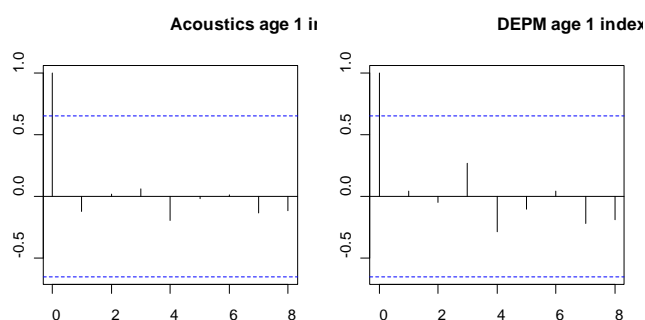


Figure 2. Autocorrelation analysis of biomass indices for age 1 for the period 2000–2009 (right).

Model fits

Standardised residuals were plotted against years to check patterns in residuals (Figure 3a-c). No obvious patterns occurred apart from an autocorrelation of standardised residuals in the final 5–8 years for total biomass estimates; in particular for the DEPM survey indices. Quantile-quantile plots (Figure 3d-f) showed that residuals for total biomass indices were approximately lognormally distributed as assumed in the model while residuals for recruit biomass did not follow the assumed lognormal distribution as they do not lie on the diagonal line as expected.

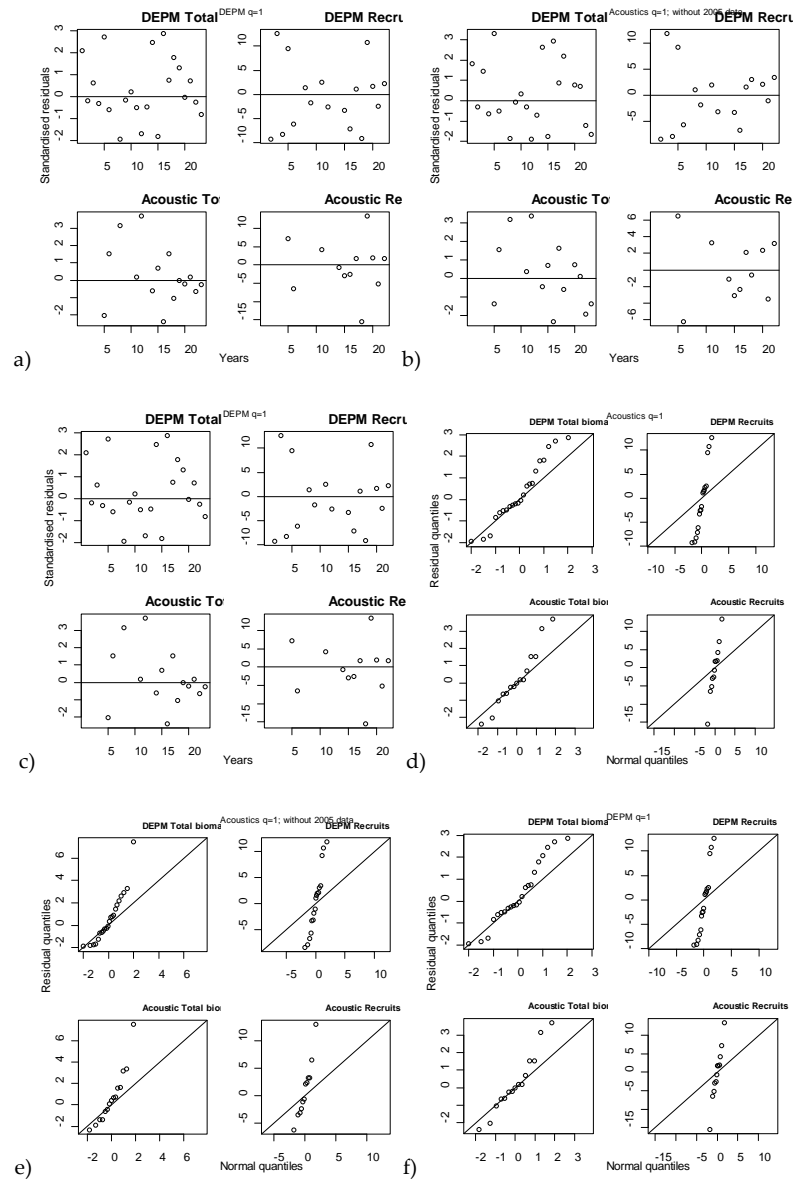


Figure 3. Residual plots. Standardised residuals and quantile-quantile plots by survey series. a) & d) $q_b^{\text{acoustic}}=1$; b) & e) $q_b^{\text{acoustic}}=1$, without 2005 data; c) & f) $q_b^{\text{DEPM}}=1$.

Relative stock estimates

The Brem model provides relative stock estimates whose absolute level is conditioned by the assumptions made for survey catchability. Setting $q_b^{\text{acoustic}}=1$ led to systematically higher biomass estimates (black continuous line in Figure 3) compared to the case than $q_b^{\text{DEPM}}=1$ (red dashed line in Figure 4). However, relative time-trends were similar. Removing data for 2005 ($q_b^{\text{acoustic}}=1$) affected all estimates for the years 2005–2009 (green dotted line in Figure 4). In particular, recruit estimates were increased and total biomass estimates somewhat decreased in recent years. Total biomass estimates including CVs are provided in table 1. Setting $q_b^{\text{acoustic}}=1$ and including data for the year 2005 provided generally the most precise (smallest CV) biomass estimates, though not for the final years.

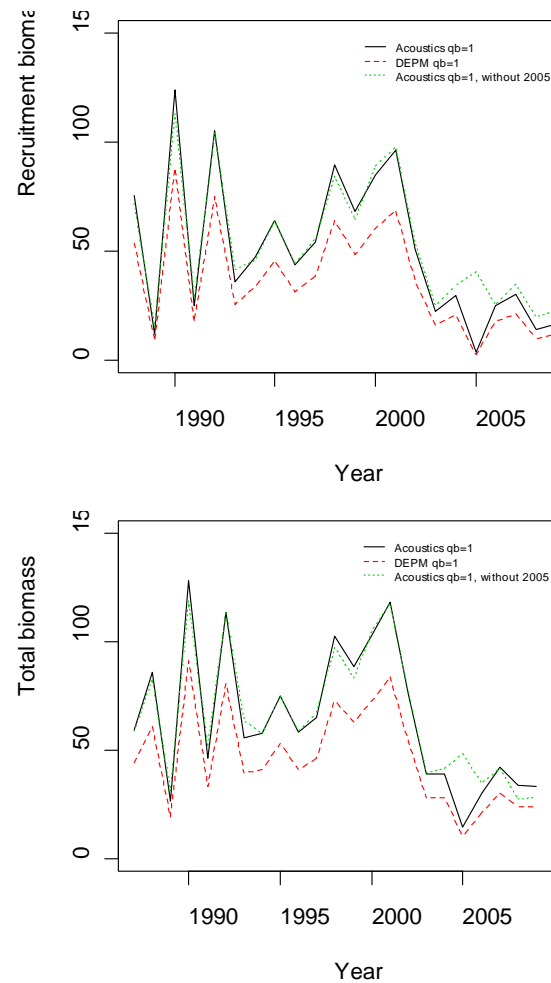


Figure 4. Model estimates for anchovy recruit biomass (age 1) and total biomass in the Bay of Biscay using acoustic and DEPM biomass indices. The black and red lines refer to models with different hypothesis on survey catchability. The green line was obtained when data for 2005 was removed.

Table 1. Relative total biomass estimates using Brem model and two survey time-series.

Year	QBACOUSTIC=1		QBACOUSTIC=1 & WITHOUT 2005 DATA		QBDEPMC=1	
	Total biomass	CV	Total Biomass	CV	Total Biomass	CV
1987	59.26	0.28	58.62	0.33	44.30	0.26
1988	85.73	0.28	82.91	0.44	61.08	0.27
1989	26.56	0.53	30.88	0.77	18.89	0.53
1990	128.41	0.27	119.04	0.44	91.19	0.26
1991	46.51	0.46	51.35	0.66	33.03	0.46
1992	113.51	0.25	114.04	0.38	80.61	0.24
1993	55.75	0.70	63.32	0.77	39.59	0.70
1994	57.55	0.41	57.99	0.63	40.87	0.40
1995	74.75	0.32	75.00	0.49	53.08	0.31
1996	58.05	0.43	58.71	0.59	41.22	0.42
1997	65.09	0.32	66.94	0.47	46.22	0.31
1998	102.61	0.37	97.41	0.50	72.87	0.36

	QBACOUSTIC=1		QBACOUSTIC=1 & WITHOUT 2005 DATA		QBDEPMC=1	
1999	88.65	0.45	83.16	0.63	62.96	0.44
2000	102.99	0.29	105.34	0.35	73.14	0.29
2001	118.12	0.32	118.05	0.46	83.89	0.32
2002	75.99	0.48	76.40	0.64	53.96	0.48
2003	39.12	0.66	39.65	0.75	27.78	0.65
2004	39.08	0.45	41.72	0.45	27.75	0.45
2005	14.61	1.11	48.45	0.68	10.37	1.10
2006	29.97	0.43	34.74	0.56	21.29	0.43
2007	42.21	0.46	41.55	0.39	29.97	0.46
2008	33.88	0.93	27.63	0.58	24.06	0.92
2009	33.44	1.06	27.98	0.53	23.75	1.05

2.2.1.3 Summary of WKADSAM discussion

Recruitment (age 1) biomass is modelled separately to total biomass (which includes recruitment) in order to make full use of recruitment information which for the anchovy application comes from acoustic and DEPM survey estimates. For Biscay anchovy, the population dynamics is driven by recruitment, with a comparatively small amount of the population subsequently contributing to the total biomass, hence why it is useful to model recruitment in addition to total biomass.

The model as applied to Biscay anchovy has potentially 4 estimates of catchability q (one each for the survey type – acoustic and DEPM – and one each for the stage – recruitment and total biomass), but the recruitment and total biomass q are estimated separately. This gives four q estimable parameters, but 1 is needed to be fixed (to 1) while the others are estimated. Which of these is fixed does matter because the two sets of time-series (acoustic vs. DEPM) have missing data in different places.

It may have been useful to include catch data for the period for which it was still regarded as reliable (if there was such a period), as this would have helped with the scaling, but this model was developed in the context of producing survey-only methodology, so it was not appropriate to include catch data in this case.

The two time-series for recruitment (age 1) biomass (acoustic and DEPM) showed good agreement (although there were also differences) – it was pointed out that the age 1 estimates for the two surveys were not entirely independent as the age structure from the acoustic estimates were used to partition the DEPM estimates after the surveys were completed. There was no exchange of information between the surveys for the total biomass.

Concerns were raised about the 2005 estimates because of inconsistency in q (fish might have been close in-shore that year so may have been missed). An exploratory run excluded 2005 entirely from the analysis, and resulted in changed estimates for 2005 and a re-scaling of the total biomass trajectory. However, omitting 2005 may imply that, given that recruitment is modelled as a random effect based on a log-normal distribution, the model will simply replace the 2005 data with the mean of the assumed distribution. Questions were raised about whether this was more justifiable than using the actual data.

Estimates of population trends from BREM showed more variation than the SAM model (a state-space model, which also uses the random effects concept), but this is because recruitment is very variable for Biscay anchovy, resulting in the random

walk process for biomass growth contributing less to the overall dynamics. However, the SAM model does not always produce smooth population trends – this depends on the underlying data.

An analysis of the q-q plot for the random effects in recruitment showed that the lognormal assumption for recruitment was not ideal. Estimating recruitment values each year as an alternative to modelling random effects was not tried. There was a problem with estimation the variance for the g random effect because of the highly variable recruitment (which is the other random effect). For Biscay anchovy, the two variances associated with recruitment and biomass growth were not easily identifiable; although the variance for recruitment could be always estimated, it could not always for biomass growth. This is probably indicating that the model is close to not being identifiable.

Problems with assuming a lognormal distribution for recruitment were highlighted. Essentially, the problem is one of asymmetry because strong year classes have a lot of information about their strength, but weak year classes are lost in the noise. Estimates of recruitment have distributions with thinner lower tail than the actual recruitment, so that the really weak year classes are not estimated to be as weak as they should be. In contrast, estimating the strength of strong year classes does not pose the same problem.

Given that for BREM, catch is excluded, and recruitment is not dependent on biomass, this must pose difficulties for communicating results to managers. Nevertheless, the survey does contain information about recruitment and impact of fishing.

The ease of use of the BREM model depends on the user's knowledge of ADMB code, because there is no front-end available. The available code could be used as a starting point for adapting the approach to a particular situation and data. However, there is the problem that ICES WGs do not necessarily contain the sort of people that can do this.

As with any model, the application to a new stock will always need an expert overview of what is done initially. This requirement can be relaxed once the method is ready for repeated use.

2.3 Stock Synthesis 3 (version 3.11b)

Stock Synthesis (SS) provides a statistical framework for calibration of a population dynamics model using a diversity of fishery and survey data. The following refers to version 3.11b, dated September 2010.

2.3.1 Description

Language

SS currently is compiled using ADMB version 7.0.1 using the Microsoft C++ compiler version 6.0.

Programmer / Contact Person

Dr Richard D. Methot, Jr., NOAA Fisheries – Office of Science and Technology, Northwest Fisheries Science Center, 2725 Montlake Boulevard East, Seattle, WA 98112. e-mails: Richard.Methot@noaa.gov

Distribution Limitation

The model and a graphical user interface are available from the NOAA Fisheries Stock Assessment Toolbox website: <http://nft.nefsc.noaa.gov/>. Only executable code is routinely distributed, along with manual and sample files. However, under certain circumstances, source code may be obtained from the author upon request and with agreement to certain restrictions.

Compiler Needs / Stand-Alone

SS runs as a DOS program with text-based input or can be invoked from a graphical interface (GUI). SS is compiled to run under a 32-bit Windows operating system, but has also been successfully compiled for LINUX (contact author for details). It is recommended that the computer have at least a 2.0 GHz processor and 2 GB of RAM. The GUI version requires only an operating system to run, has been written to use the Microsoft .Net framework and to support enhanced features such as screen resizing. However the GUI version does not support all features of SS. These features, particularly tag-recapture and generalized size frequency, are fully described in the user manual and can be invoked by editing the input files using any text editor. The same executable program, SS3.exe, is used in association with the GUI or directly with text files. Output is written to a set of files which can be read using a text editor. However, to facilitate visualization of the results output processors are available for Microsoft Excel and R. Up to date versions of SS code and documentation are cataloged by the United States NOAA Fisheries "Toolbox" at <http://nft.nefsc.noaa.gov/SS3.html#About>. Code and instructions for the R output processor are available at <http://code.google.com/p/r4ss/>.

Purpose

Stock Synthesis provides a statistical framework for calibration of a population dynamics model using a diversity of fishery and survey data. Stock Synthesis is designed to accommodate age-structured, size-structured, and age-aggregated data within a population model that can include multiple stock subareas. Thus it is most similar to A-SCALA (Maunders and Watters, 2003); Multifan (Fournier *et al.*, 1990); Multifan-CL (Fournier, Hampton and Siebert, 1998); Stock Synthesis (Methot 2000) and CASAL (Bull, *et al.*, 2004) in basic structure and intent. A general feature of such models is that they tend to cast the goodness-of-fit to the model in terms of quantities that retain the characteristics of the raw data. For example, age composition data that is affected by ageing imprecision is incorporated by building a submodel of the ageing imprecision process, rather than to pre-process the ageing data in an attempt to remove the effect of ageing imprecision. By building all relevant processes into the model and estimating goodness-of-fit in terms of the original data, we are more confident that the final estimates of model precision will include the relevant sources of variance. SS is designed to provide a highly scalable approach that is not critically dependent on having particular types of data. This allows SS to analyse long time-series that extend from data-weak historical periods into data-rich contemporary periods. SS also directly incorporates stock density-dependence by modelling annual recruitment as deviations from an estimated spawner-recruitment relationship. This allows SS to internally calculate F_{msy} and other benchmark quantities, and to use these quantities in forecasts of potential yield and future stock conditions. This complete integration allows SS to produce confidence intervals on these quantities. Examples of routine outputs include the probability that a proposed TAC will produce overfishing next year, and the probability that a proposed harvest policy would leave the stock above a specified biomass threshold 5 years in the future.

Description

The overall SS2 model is subdivided into three submodels. First is the population dynamics submodel. Here the basic abundance, mortality and growth functions operate to create a synthetic representation of the true population. Second is the observation submodel. This contains the processes and filters designed to derive expected values for the various types of data. For example, survey catchability relates population abundance to the units in which survey cpue is measured; an ageing imprecision matrix transforms the estimated sampled numbers-at-age into an estimate of the proportions recorded in each otolith ring count. Third is the statistical submodel that quantifies the magnitude of difference between the various types of data and their expected values and employs an algorithm to search for the set of parameters that maximizes the goodness-of-fit. An additional model layer is the estimation of management quantities, such as a short-term forecast of the catch level that would implement a specified fishing mortality policy. By integrating this management layer into the overall model, the variance of the estimated parameters can be propagated to the management quantities, thus facilitating a description of the risk of various possible management scenarios.

The complexity of the population submodel should be considered relative to the complexity of the data and observation submodel. For example, if only biomass-based cpue data are available, it is simplest to cast the population submodel as a simple biomass-dynamics model such as the delay-difference model (reference). However, with integrated analysis it is possible to build a more complex, age-structured population submodel that collapses to the simple biomass level in the observation submodel. If the various mortality, growth and selectivity parameters necessary in the more complex model are fixed at levels that mimic the inherent assumptions of the simple biomass dynamics model, then both models produce identical results. The advantage of the more complex internal model is that it is primed for a richer array of sensitivity testing and immediate incorporation of more detailed data as these data become available.

The model to be presented here is primarily designed for a particular, although not overly restrictive, set of circumstances and data. The target species are groundfish that are harvested by multiple distinct fleets and for which there commonly are fishery-independent surveys to provide a time-series of relative abundance. Some age and length composition data are available from both the fishery and survey, but they are intermittent, often based on small sample sizes, and the age data are influenced by a substantial degree of ageing imprecision. Tagging data are not available for these species and analysis of tagging data has not been built into the observation submodel.

Program Inputs

Many types of data may be input to SS, but no one data type is required for a model to run. Some parameters are required while others are conditional on the model configuration, depending on such options as multiple areas, growth patterns, etc. Please see the user's manual for a complete description of the inputs and a discussion of the appropriate usage.

The potential data inputs include:

- Dimensions (years, ages, N fleets, N surveys, etc.)
- Fleet and survey names, timing. Etc.
- Catch biomass

- Discards
- Mean body weight
- Length composition set-up
- Length composition
- Age composition set-up
- Ageing imprecision definitions
- Age composition
- Mean length or bodyweight-at-age
- Generalized size composition (e.g. weight frequency)
- Tag-recapture
- Stock composition
- Environmental data

In addition, there are required and optional parameter inputs. Optional inputs are required for more complex model formulation (e.g. multiple growth patterns, sub-morphs, areas). The correct specification of these parameters is complex, but is fully described in the user's manual.

- Number of growth patterns and sub-morphs
- Design matrix for assignment of recruitment to area/season/growth pattern
- Design matrix for movement between areas
- Number of and definition of time blocks that can be used for time-varying parameters
- Specifications for mortality, growth and fecundity
- Natural mortality and growth parameters for each gender x growth pattern
- Maturity, fecundity and weight-length for each gender
- Recruitment distribution parameters for each area, season, growth pattern
- Cohort growth deviation
- Environmental link parameters for any biological parameters that use a link
- Time-varying setup for any biological parameters that use blocks
- Seasonal effects on biology parameters
- Phase for any biological parameters that use annual deviations
- Spawner-Recruitment parameters
- Recruitment deviations
- Method for calculating fishing mortality (F)
- Initial equilibrium F for each fleet
- Catchability (Q) setup for each fleet and survey
- Catchability parameters
- Length selectivity, retention, discard mortality setup for each fleet and survey
- Age selectivity setup for each fleet and survey
- Parameters for length selectivity, retention, discard mortality for each fleet and survey

- Parameters for age selectivity for each fleet and survey
- Environmental link parameters for any selectivity/retention parameters that use a link
- Time-varying setup for any selectivity/retention parameters
- Tag-recapture parameters
- Variance adjustments
- Error structure for discard and mean body weight
- Controls for weighting likelihood components (lambdas)

Program Outputs

The major sections of the primary output file (Report.sso) are listed below. Each section has an associated label. Additional output files include a database of the model results for composition data (compreport.sso) and the covariance between all pairs of estimated parameters and key derived quantities (covar.sso). The Excel spreadsheet ss3-output.xlsx reads the files report.sso and compreport.sso, and searches for labels in the first column. The data are then automatically copied into specific worksheets for detailed display. Similar capability is also available using R routines (included in the software catalogue) and has been included in the GUI.

A summary of the major sections of the output file is as follows. A detailed description of each output can be found in the attached user's manual.

- SS version number with date compiled.
- User comments transferred automatically from the input files.
- List of keywords used in searching for output sections.
- List of fishing fleet and survey names.
- Final values of the negative log(likelihood) and the components associated with each data type and fleet or survey.
- The matrix of input variance adjustments.
- Parameters.
 - For the estimated parameters, the output includes: the count of parameters, an internally generated label, the value, the active count (a count of the parameters in the same order they appear in the ss3.cor file.), phase, minimum, maximum, initial estimate, the prior, type of prior, standard deviation of the prior, the likelihood of the prior, the standard deviation of parameter as calculated from inverse Hessian, the status (e.g. near bound) and the value of prior penalty if parameter was near bound. Please see the user's manual for a complete description.
- Derived quantities
 - This section outputs the options selected from the starter.ss and forecast.ss input files, then the time-series of derived quantities, with standard deviation of estimates. The output includes such quantities as spawning biomass, recruitment, SPR ratio, F ratio, B ratio, forecast catch (as a target level), forecast catch as a limit level (OFL). There are also additional outputs (e.g. Selex_std, Grow_std, NatAge_std) which are explained in the attached user's manual.
- Biological parameters, by year, after any time-varying adjustments.

- Size selectivity parameters, after by year, after any time-varying adjustments.
- Age selectivity parameters, by year, after any time-varying adjustments.
- Distribution of recruitment across growth patterns, genders, birth seasons, and areas in the end year of the model.
- Morph indexing
 - This block shows the internal index values for various quantities. Please see the user's manual for a complete description.
- Size-frequency translation
 - If the generalized size frequency approach is used, this block shows the translation probabilities between population length bins and the units of the defined size frequency method. If the method uses body weight as the accumulator, then output is in corresponding units.
- Movement rates between areas in a multi-area model.
- Exploitation: The time-series of the selected F_{std} unit and the F multiplier for each fleet in terms of harvest rate (if Pope's approximation is used) or fully selected F .
- Time-series
 - The time-series of abundance, recruitment and catch for each of the areas. Output quantities include summary biomass and summary numbers for each gender and growth pattern. For each fishing fleet, the output includes: encountered (e.g. selected) catch biomass, dead catch biomass, retained catch biomass, the same three quantities in terms of numbers, the observed catch and the fully selected F multiplier.
 - The final column shows the spawning biomass as calculated with the start year's fecundity-at-age. If there are time-varying life-history parameters, this column will show the impact of these changes on the calculation of spawning biomass in comparison to the spawning biomass calculated with the current year's life history.
- Spawning potential ratio (SPR) time-series
 - This section reports on the yield-per-recruit and biomass per recruit calculations according to the current year's life-history, fishery selectivity and fishing intensity. It is annual, so accumulates the effects across seasons and areas. The report includes the forecast period. The level of recruitment used for the calculations is the virgin recruitment level (R_0). The details of the outputs contained in this section can be found in the attached user's manual.
- Spawner output and recruitment
 - This section includes the estimated recruitment according to the spawner-recruit curve, the adjusted recruitment according to the input environmental conditions for that year, the bias-adjusted expected mean recruitment, and the predicted recruitment used in the model after adjusting for the year specific recruitment deviation, and additional outputs. Please refer to the user's manual for a detailed description of the outputs contained in this section, and the use of these outputs by the model.
- Details on each index data point.

- Values include input observed values, expected values, expected vulnerable biomass, catchability, and likelihood contribution.
- General information on each index time-series and associated parameters- Observed and expected values for the amount (or fraction) discarded.
- Observed and expected values for the mean body weight.
- Goodness of fit to the length compositions.
 - The input and output levels of effective sample size are shown as a guide to adjusting the input levels to better match the model's ability to replicate these observations.
- Goodness of fit to the age compositions. Same format as the length composition section.
- Goodness of fit to the size compositions. Same format as the length composition section. Used for the generalized size composition summary.
- Length selectivity and other length specific quantities for each fishery and survey.
- Age selectivity and time-series of length-based variables converted to age. Selectivity, fecundity, and mean body weight are all converted from functions of length to functions of age using the modelled distribution of size-at-age in each year.
- The input values of environmental data are echoed in the output file.
- Numbers at age (in thousands of fish) are shown for each cohort (combination of birth year, gender, area, etc.) tracked in the model.
- Numbers at length (in thousands of fish) are shown for each cohort tracked in the model.
- Catch at age is shown for each combination of cohort and fleet. Not by area because each fleet operates in only one area.
- Biology: The first biology section shows the length-specific quantities in the ending year of the time-series only. The derived quantity spawn is the product of female body weight, maturity and fecundity per weight. The second section shows natural mortality.
- Growth parameters: the biology parameters, and associated derived quantities.
- Biology at age: the derived size-at-age and other quantities calculated in the end year of the model.
- Mean body weight (beginning): the time-series of mean body weight for each morph. Values shown are for the beginning of each season of each year.
- Mean size time-series: the time-series of mean length-at-age for each morph. At the bottom is the average mean size as the weighted average across all morphs for each gender.
- Age-Length Key: the calculated distribution in each population length bin at each age for each growth morph at the midpoint of each season in the ending year.
- Age-Age Key: the calculated distribution of observed ages for each true age for each of the defined ageing keys.
- Selectivity database: the selectivities organized as a database, rather than as a set of vectors.

- Spawning Biomass Reports (1 and 2)
 - This section contains annual total spawning biomass, then numbers-at-age at the beginning of each year for each bio pattern and gender as summed over sub-morphs and areas. Next, Z-at-age is reported simply as $\ln(N_{t+1,a+1} / N_{t,a})$. Then the Report_1 section loops back through the time-series with all F values set to zero so that a dynamic Bzero, N-at-age, and M-at-age can be reported. The difference between Report_1 and Report_2 can be used to create an aggregate F-at-age.
- Length/Age/Size composition database:
 - This is reported to a separate file, compreport.sso, and contains the length composition, age composition, and mean size-at-age observed and expected values. It is arranged in a database format, rather than an array of vectors. Software to filter the output allows display of subsets of the database.
- Variance and Covariance of parameters and derived quantities:
 - This is reported in a separate file, covar.sso.

Diagnostics

Various model diagnostics are contained throughout the output files. These include the following:

- Likelihood contributions from each data type for each fleet or survey
- Likelihood contributions from each data point
- Variance for each estimated parameter and key derived quantities and covariance between them
- Information on which parameters are hitting bounds or not moving from initial values
- Effective sample sizes for composition data
- Detailed warning for incorrect or inadvisable model setups
- Optional output of parameter values at every step of estimation.
- Optionally: retrospective analysis, likelihood profiling, MCMC, and simulating data with same structure as input data for testing estimability of parameters

Other Features

- Parameters have a wide array of options, including options for phased estimation, constraints, Bayesian priors, temporal variation, and other options. See user manual for more details on these options.
- Bayesian posteriors for parameters can be calculated using MCMC.
- Population status relative to a variety of benchmark quantities can be calculated.
- Forecasts can be conducted under a variety of harvest policies.

Ease of Use

- Input values are echoed to the file "echoinput.sso" so the user can see exactly how SS is reading and interpreting the input files.
- SS now produces four files (*.ss_new) that mirror the input files and provide text to describe the range of options available. This text will make it

easier for users to take existing files and modify them to meet the conditions they which to invoke.

- All new features and several older features now have conditional inputs. This means that the user no longer needs to provide placeholder inputs for features they are not using. In addition, commented out placeholders for unused features still are output to the `ss_new` file so their syntax is visible for future use.
- Some model features are now bundled into collections of advanced features. If the user sets an advanced feature flag to a value of 0, then there is no reading of that set of advanced features, SS assigns appropriate values or null values as necessary, and these assigned values appear as commented out output in the `ss_new` file.
- Many more situations that are illegal or not optimal are identified and a warning is output to `warning.sso`. The total number of warnings is displayed to the screen at the end of the run.
- SS internally creates a text string to label each parameter. These labels are used to identify the parameters in the new control file (`control.ss_new`), in the report file (`Report.SSO`), and in the covariance matrix output (`COVAR.SSO`) which is now output as a user-friendly database rather than a matrix.
- Comments can be read and stored if they are placed before the first valid input line in each input file and if each comment line begins with the characters `#C`. Blank lines and lines starting with just a `#` are ignored. These comment lines are written out at the top of `report.sso` and at the top of the `*.ss_new` files.

Code fixes and internal revision

These are described in detail in the user's manual.

History of Method Peer Review

SS has been used for dozens of stock assessments around the world. The area of highest use is on the US Pacific Coast. Numerous stock assessments conducted by NMFS scientists at the Northwest and Southwest Fisheries Science centres using SS have been reviewed by a stock assessment review (STAR) panel which includes independent CIE reviewers. These assessments are then reviewed by the Scientific and Statistical Committee of the Pacific Fishery Management Council.

Steps Taken By Programmer for Validation

An example application was created to demonstrate the model's basic capabilities. This was a simple test in that the growth, natural mortality and form of selectivity patterns were set to be identical with that in the model used to generate the simulation data. Nevertheless, the demonstration shows the ability of the model to correctly deal with variability in input data. The dataset spanned 1971- 2001. A single fishery was implemented with a constant logistic selectivity pattern over time and with age and size composition data each year. Natural mortality was fixed at 0.1 and the accumulator age in the population was set at 40 years. A triennial fishery-independent survey was developed and used during 1977–2001 with associated age and size composition. An annual recruitment index for each year 1990–2001 was also used. The recruitment followed a Beverton–Holt spawner-recruitment relationship with steepness of 0.7 and standard deviation of residuals equal to 0.8.

In order to demonstrate the ability of the model to achieve size-specific survivorship the following morph structure was adopted: there was one male growth morph and five female growth morphs; the male growth morph and the middle female growth morph had identical mean size-at-age; the male growth curve had a broad variability of size-at-age and each female growth morph had a narrower variability, but the set of female growth morphs together produced an unfished size-at-age distribution that was identical with the male distribution. Twenty parametric bootstrap datasets were generated by SS from one population realization. Male size-at-age is one broad morph; female size-at-age is composed of five narrower morphs.

The results of the simulation study demonstrate the ability of SS to correctly deal with variability in the simulated input data, and return the expected (“true”) result. The true spawning-stock biomass trajectory and the mean and median estimated spawning-stock biomass trajectory of the twenty bootstrapped datasets are nearly identical (Figure 1).

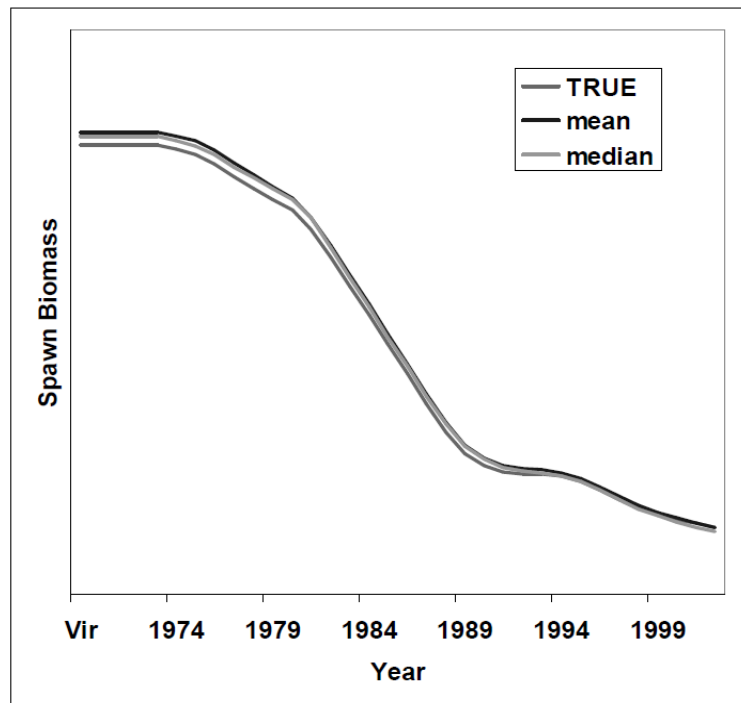


Figure 1. The average estimate of spawning biomass among the fits to the 20 datasets is nearly identical with the true level.

Likewise, the true annual recruitment strongly resembles the mean estimated from the twenty bootstrapped datasets. Although, it is evident that the average estimate of recruitment slightly undershoots the true range of variability, as expected for a model with imperfect data (Figure 2).

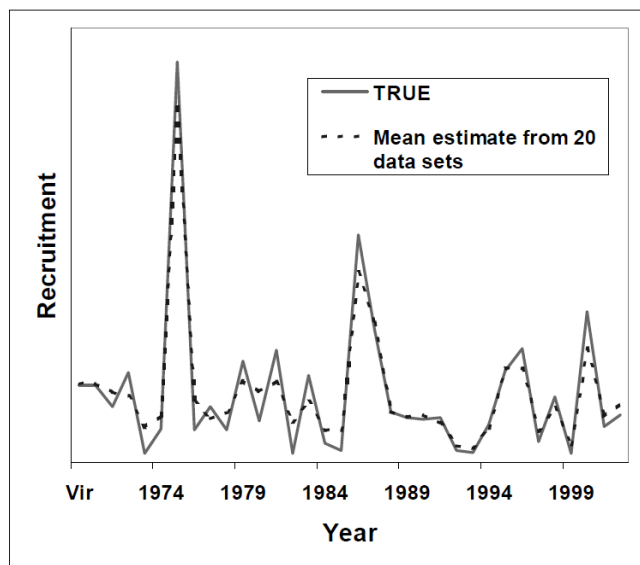


Figure 2. The true recruitment and the average estimated recruitment from 20 bootstrapped data-sets.

Finally, the parametric estimate of variance in spawning-stock biomass was estimated from the inverse Hessian within ADMB and compared to the variability among the fits to the 20 bootstrapped dataset. The results were quite similar, suggesting that SS properly characterized the variability in the simulated data (Figure 3).

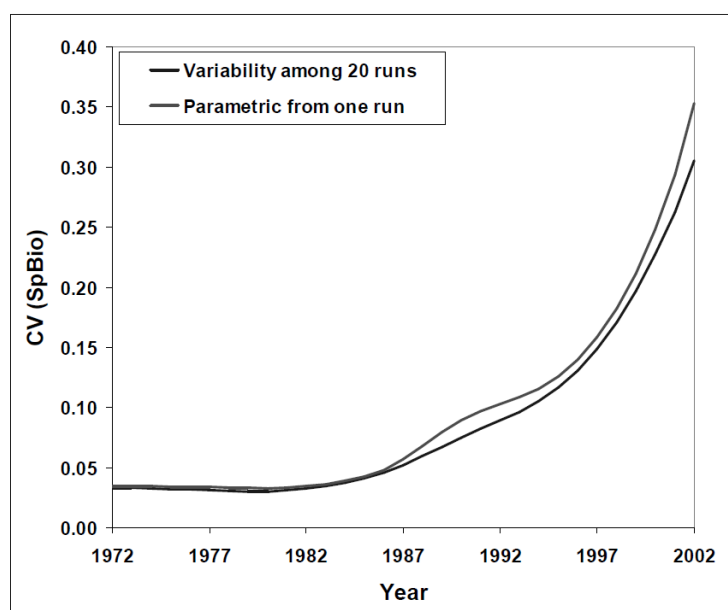


Figure 3. The parametric estimate of variance in SSB estimated from ADMB's inverse Hessian is nearly identical with the variability among the fits to the 20 datasets.

Tests Conducted By Others

Numerous tests have been conducted using this model. Those published in peer reviewed literature include Yin and Sampson (2004), which reached the conclusion that "For all the output variables examined the estimates appeared to be median-unbiased", wq and Schirripa *et al.* (2009) which focused on incorporating climate data,

but provided an additional check of the ability of the model to estimate parameters using simulated data.

Various ongoing research projects have determined that SS is generally capable of estimating parameters used to simulate data. These include the work of Maunder *et al.* (2009) and separate projects being conducted by Ian Taylor, Tommy Garrison, and Chantel Wetzel, all associated with the University of Washington.

2.3.2 Summary of WKADSAM discussion

Rick Methot presented the Stock-Synthesis (SS) stock assessment method. The key discussion points raised following the presentation were:

- In the first simple example, it was clarified that landings and discards were combined for the simple model.
- In the process example, there were large differences in the size of age 4's kept in the catch vs. discarded. It was mentioned that there was size-based age sampling.
- With time-varying growth, if fish are greater than L_{inf} (when in changes) then the fish are modelled to not grow any further.
- In the Grouper example, it was questioned if the switch from female to male was density-dependent? The response was no, but they are close to doing this.
- Propagation of uncertainty. SS includes uncertainty in forecasted quantities.
- Clarified that F is for total catch, and in the model the catch is decomposed into retained and discarded, and the retained model catch is fitted to observed catch.
- Morphs and sub-morphs. Morphs are a group of fish with different biological characteristic. Sub-morphs are nested within morphs.
- The index error must be supplied to SS, but there is an option to estimate extra variance for indices.
- A difficult issue is: how do you know when to stop modelling the processes that generate the data? R. Method responded that he would like to see assessments evolve so that the data decides model structure. Also, models need to become more spatially explicit.
- There was discussion on how easy SS could be implemented in the ICES context. The model is very flexible, but there is a penalty attached to this. Some ICES stock assessment scientists only do stock assessment 2 weeks a year, and in this case it is difficult to build and maintain competency with the method. Also, the model flexibility means there are many subjective decisions to make and the assessment can become subjective. R Method suggested that stock assessments should have research models and production assessment models. The assessment model complexity issue is very real. SS should start off as a research model and not an assessment production model. What we learn about processes from SS is important when understanding other models.
- It was noted that there is a connection between the 'stiffness' of models and bias and retrospective problems. However, flexible models can explain residual patterns in different ways. How can we deal with this? R. Method

agreed. Different communities treat this differently (i.e. assume M, or flat selectivity), and too many fixes may mean the data cannot be fitted.

- Users with agendas can use flexibility to suit their agenda. There is no good answer for this. Also, review panels are also an issue; they change things and sometimes don't have a good understanding of the changes.
- SS is complex, but a particular assessment does not have to be complex. Many of SS options can be turned off. R. Method agreed, and suggested users take a hierarchical approach, and start simple first.
- SS can do a length-based analysis, but is it better to just code this up yourself. Which is better? Scalability is important. Paring of simple and more detailed models is important.
- There is a danger of over-fitting with a detailed model.
- We have gone to more complex models over the years to better account for uncertainty. But it is difficult to account for uncertainty in more complex models. Should we only build models that we can accurately account for uncertainty for (i.e. using MCMC, etc)? R. Method suggested that this changes over time, and it is a trade-off.
- One desirable feature of simple/stiff models is that you are consistent over time. With flexible models there can be less consistency.
- Can SS be used to explore source of retrospective patterns? R. Method replied that the model allows user to cover the sources, and this should be done.

2.4 MULTIFAN-CL

2.4.1 Description

No summary of the model package was provided by the presenter.

Rapporteur's summary of presentation

The presenter started by saying he was giving just a flavour of Multi-fan and what it was designed for and what it can and cannot do. It has been designed for highly migratory species. He will talk about some of the auxiliary software developed for projections and such.

Some of the main challenges are because of large spatial scales, incomplete mixing, and often migratory behaviour. They need to take in to account the effects of spatially explicit fishing in different areas with often different fleet practices that have changed over time. One of the first problems is the lack of catch-at-age data because they are difficult to age. There are strong seasonal patterns, because many of the fish are tropical and only go into far southern and far northern latitudes in specific parts of the year. Catch-reporting by fleet varies, and it could be in weight or length. They have individual weights for some of the higher value fish. Some species have individual weights for up to 95% of the data. Some fisheries have no data at all. There are no surveys, so tagging data are the main fishery-independent data source. There is high dimensionality in the model because of the diverse data sources.

Multi-fan CL is a combination of two historical approaches, the original Fournier et al MULTIFAN and the Fournier and Archibald approach, which gives MULTIFAN-CL in Fournier et al. 1998. We are still using it 12 years later, and it is the same basic structure.

The main features include: It uses catch and effort data separately, not cpue. This helps accommodate missing values for when you have one and not the other. It can include temporal CVs for catch/effort data similar to SS. Catch can be in weight or numbers, but not different in same fishery. The catch can be either known exactly or with assigned CVs. Length and/or weight frequency data can be used. Only conventional tag data used now, but they are working toward using archival tag data for spatially explicit models.

An example of length frequency data were shown, showing cohorts moving through time. The red lines are different cohort ages.

Tagging data were shown for skipjack tuna. The tagging data are used to help estimate both movement and exploitation rate, but not used for growth.

One thing he missed is that it is an age structured model, but uses lengths to estimate those ages. The population dynamics include multi-region and many different time-steps. Seasonal and age-dependent movement is possible but not currently used. One thing to explore is whether movement is time-invariant. What we find is that tropical tunas have different movements because of events like El Niño. One thing they are considering doing is adding covariates for time-dependent movement. Growth is an example of using different time-steps. They also take advantage of the ability to use different growth because of age, such as using LVB for older and Schnute-Richards for early ages. Non-equilibrium initial conditions are possible by estimating an exploitation parameter prior to catch data. A B-H spawner-recruitment curve or independent recruitment could be used. An environmental recruitment index could be used.

An example of spatial structure was shown where maps showed structure of bigeye tuna, which is divided into six regions. The upper right quadrant showed regional movement patterns. Our plotting function gives different thicknesses, but they do not mean anything here because all movement parameters are the same at age in this model. The estimates of movement are one of the more interesting parts of the assessment because of the political implications. The figure in the bottom right quadrant showed estimates of total biomass over time by region. This can help look at ideas like range contraction. It is easier to estimate movement w/ tagging data, but northern areas can be estimated for movement because they disappear in some seasons. The question was asked whether the movement rates were the same for all ages. In this assessment they do not vary by age-classes but they could. The question was asked how the model knows the fish are not there if there is not any fishing. The presenter said there is some fishing all year. They could be constant and age-dependent and maybe changed in the future. In MF there are two ways that a fish could be in another region, move there or recruit there. This is confounded in the model because the process could happen either way, so the model does not know this. There are some interesting things that could be done like examining localized spawner-recruit events. We know that tropical tunas spawn continuously given that conditions are favourable. The pie-charts on the figure on the left present catch by gear type. Some fisheries fish when they are 12 inches, intermediate, or large.

The presenter discussed some biological processes. There is some different variation in age-at length in the early years where it departs from LVB. They model natural mortality as a U-shaped distribution quarterly. The large amount of tagging data allows quite consistent estimates of M at age. The estimates range from 0.8 to 0.2 at age 10 and back to 0.6 at age 15.

Fisheries inhabit discrete regions and F is separable into selectivity and q . Two catch equation options are to estimate Baranov instantaneous rates or simpler approximation of $C=FN$. Selectivity is flexible for length or age based versions and can be a functional form, cubic splines, or other additional types and constraints are possible. Catchability can be seasonal and/or time varying. One of the major features is the ability to link catchability and selectivity across fisheries. The way that you link fisheries together with a nested hierarchical structure such as shared selectivity and q , and are able to relax that structure where more information is available. At the start of each year, you are back at the same distribution of length-at-age.

The presenter described some fisheries in the assessment. The figure on the left showed lines that are changing catchability over time. One standardized fishery has constant q because it is standardized, while others are non-standardized and q is allowed to drift around, because they don't believe that they are indicative of abundance. In the purse-seine fishery, the model says that there have been large changes in the efficiency. He shows plots of the effort deviates. Selectivity curves show different shapes from asymptotic to descending limb curves.

Model diagnostics include looking at observed vs. predicted tag-recaptures. Residual plots were shown for looking at fit to catch-at-length data. The data are complex. From 1950–2010 using lengths from 50–200 cm, blue circles mean that there is more catch in the data than the model predicts at length. It shifted from not much pattern to too many small fish, to too many big fish, and back to too many small fish. When the model sees these big trends, it probably is not estimating recruitment well because it is based on length. So the separability assumption seems to not be working well in this case as selectivity may be changing. This could be spatial or gear-related, or the sampling is not-representative. It could also be temporal changes in the movement patterns. These all might show up in the recruitment patterns. It can be expected that any dataset over sixty years, data collection changes.

For 10 years, tuna assessments did not have a management forum. Now that there is a tuna commission, this has changed. So now we estimate MSY-related quantities. Another new thing is a dynamic MSY that accounts for fleets changing characteristics or changes in productivity in the environment. Estimation of depletion-based reference points through “no fishing” analyses can be used to show the effect of a particular fleet/country on the stock. They can do deterministic projections with mixed catch and effort limits to examine complicated quota types. Now they are incorporating stochastic error from current estimation error and future recruitment projection errors.

He showed some management plots which are the typical phase-plane plot with colours. It plots F/F_{msy} vs. B/B_{msy} . This summarizes the management trajectory over time. These may or may not include movement data. A second plot is the fishery impact plot, which shows that, for example, a value of 80% means 80% of biomass is being removed by fishing. This is a good way to compare gear types or countries impact the stock. These are from removing fishing from the model in various parts. We see significant changes in recruitment over time so we use a dynamic baseline.

In terms of uncertainty, they generally present three methods. They use the hessian matrix to show approximate confidence intervals of biomass trajectories. In the second plot, they use the likelihood profile for the same phase-plane plot shown on the previous slide. MCMC is not really possible for these models because of their dimensionality, but does not matter given the range of uncertainty between models. Instead they profile over individual models, so they show results plotted over 128 models.

When they look at different structures such as this, it is much larger uncertainty. The question was asked whether we should present that to managers. The link between the assessment results and management action is weak. There are no fine-scale decision rules. The question was asked whether they assign probabilities to each scenario. Not yet, but that might be a good thing to do in the future. CN: That is an example of presenting something based on what the managers will do. Could be used be treacherously. What does it tell you about the stock? This particular example shows a profiling of steepness, and how important it is.

Logistics of the model include: It is usable in windows or Linux executables. It is best run in Linux. When running Multi-fan CL, the estimation in phases is critical for convergence. It is necessary to use various penalties during early parts of estimation. These penalties are not only user-defined; some penalties are internal to the model where required.

The auxiliary software for MultiFan CL includes MUFDAGER for accessing the database. R is used for all post-processing and R makes calls to Multifan. Condor is a high thorough put computing environment to do multiple jobs across the network. Multifan-CL viewer is a Java based application for quick visualization of results.

Overall the pros and cons:

One pro of the model is built by the Creator and He just knows. It is custom made specifically for highly migratory species. It is developed by a programmer and runs on multiple platforms. There are constant development which is conducted to meet the needs of both assessment scientists and managers. The cons are that it is well over ten years old and does not currently accommodate critical groundfish features like age data. It does not necessarily handle length based models. The code is starting to get a little difficult to follow as additional modules have evolved.

2.4.2 Summary of WKADSAM discussion

A question was asked about using ages and the presenter stated that it was indeed an age-structured model, but it predicts lengths.

There is a website for MULTIFAN-CL that has examples, code, software, etc.

He then showed TUMAS management software. They are developing software that allows managers to test general scenarios. It allows users to easily modify some fisheries, run the projection model, and shows biomass by region. The user can click a button and get estimates quickly into Excel. Catch data also can be extracted by fleet to do bioeconomic modelling.

It was asked if TUMAS projections were deterministic. It was stated that they are. Model speed was discussed. One option to speed up the model is to reduce bin size. Anders asked about CASAL whether it used ADMB. It does not. Another question was asked whether Dave Fournier was needed to maintain Multi-fan CL. The Multifan CL users are planning contingencies for that.

Ray asked a general question about the variance across models vs. within. Usually you are looking at a few candidates, not 128. In this case the MCMCs can be quite useful to look at uncertainty across models. In general though, which applies to all highly parameterized models, with all these parameters and phases, there are a lot of decisions need to be made. At the end of the day what to you do to ensure convergence to a global solution? We have mixed up the phasing and penalties to force it to start in different places. He jitters the starting values to make sure he gets to the same place. Likelihood profiling can help you find a local minimum, but has found non-

global minima have been inconsequential in terms of management quantities. Picking the phases is usually done by importance of the parameters. It is troubling that these supermodels are very difficult to check because of the run time. A good indication of convergence problems is when you have obviously wrong results, not when they are close. There are ways to address convergence in simpler models that are unavailable in big models which make it easier to identify why a model ends up in a strange place. In SS, there are features to make it focus on one part of the dataset then relax those assumptions in later phases.

It was pointed out that the model has 3000+ parameters and it was asked what comprised the bulk of the parameters. They are mainly effort deviates and time varying catchabilities. They are closer to random effects than fixed effects because they are constrained. When you fix the catch, you iteratively solve the Baranov equation instead, and it takes just as long. The variance is about the same in the derived SSB whether you estimate F deviations or not. The catch-condition case, the implied Fs get high, the gradient surface gets to stiff and makes it slower. When F is low they are virtually equivalent. A recent implementation is to condition F early on then estimate F later on. The presentation was concluded.

2.5 CASAL

2.5.1 Description

2.5.1.1 To be written.

Rapporteur's summary of presentation

General statistical assessment tool for New Zealand specific stocks but similar to SS3. Bayesian assessment required for management advice in NZ, MCMC completed after model sensitivities based on MPD results. If model is length based it can only use length data, if model is age based you can use both age and length comps. Review of a few basic model setups. Model has settings to enable it to mimic Coleraine, and comparisons of the two models have produced very similar results. Detailed user manual is available, more so than for many other models. Error check in model setup will tell you if something is lacking in the input files. There is one modelling framework (the partition) that is used for very simple to complex models, format of observation and estimation code blocks makes it easy to run model sensitivities. With CASAL the standardized input files make it easy to understand someone else's assessment and to run it if necessary. There is a CASAL2 in the making; the scope of this is not yet determined; but, for example, it might be more spatially explicit than the current version, as managers are interested in closed areas impacts.

2.5.2 Summary of WKADSAM discussion

- Who is developing code? About 4–5 people work on this. Using a single tool has increased the skills and productivity of modelling group because everyone understands the tool. Someone else can run an assessment that they did not develop.
- CASAL is now free software. Believe the source code is not available publicly.
- What won't the model do? It will not do VPA or production models, it will only do what the model structure (the partition) will allow.
- What problems did CASAL fix? It has improved the assessment process in that all assessment scientists can use it, for a wide variety of assessments,

as it was not written for a specific problem. The over arching goal was to have a modelling framework that can be used for a range of problems of varying complexity. The model has been modified as needed based on issues arising in NZ assessments. In this respect, CASAL has been very successful.

- Can this be used for multispecies assessments? Not explicitly. It can handle multiple stocks, but not multiple species. Does CASAL include predation mortality? No. One issue here is how to handle catch – is catch for all stocks or for each stock? It would be nice to share effort time-series across multiple stocks.
- Is it mandatory to use priors? Yes, you need to provide bounds and priors. In many cases the prior is uniform.
- How long are input files? Depends upon complexity of assessment and amount of data. It could be anything from 2–20 pages. Much of this will be the input data. In simple cases the files are short, because you don't specify things that are not used.
- There is an R library associated with CASAL for reading outputs.

2.6 TINSS

2.6.1 Description

TINSS is an age-structured model that is parameterized from a management-oriented approach (Martell *et al.*, 2008). The leading parameters are MSY and F_{msy} , from which the population parameters B_0 and steepness are derived given age schedule information on selectivity, growth, maturity and natural mortality. The model is fit to data on relative abundance, age-composition and jointly estimates variance components for process errors and observation errors. Age-composition data are treated as a multivariate logistic observation and are weighted in the objective function using the conditional maximum likelihood estimate of the variance.

Nearly all of the traditional stock assessment models that jointly fit a stock–recruitment relationship to the available data estimate the unfished spawning-stock biomass (B_0) and the steepness (h) of the stock–recruitment relationship. In order to reliably estimate these parameters there must be sufficient contrast in the data to resolve the confounding of B_0 and h . That is, we must observe reductions in recruitment at low spawning stock abundance to reliably estimate quantities such as steepness. For many fish stocks, there are insufficient data to resolve this confounding and the use of informative priors is required for B_0 or steepness, or both. It is often difficult to obtain prior information for these parameters for a given stock. Moreover, the use of such priors also implies prior information about management related quantities such as MSY and F_{msy} .

TINSS differs from the traditional stock assessments in that it attempts to estimate the management related quantities MSY and F_{msy} directly. Given initial estimates of MSY , F_{msy} , and other age-specific schedule information (e.g. selectivity, maturity, and natural mortality) TINSS then calculates the corresponding values of B_0 and steepness that is required to achieve this MSY and F_{msy} . The transformation of MSY and F_{msy} to B_0 and steepness has an analytical solution; the reverse transformation does not have an analytical solution and hence it is extremely difficult to examine the implications of informative priors on B_0 and steepness on the management variables.

A simple example of this parameter transformation can be demonstrated with a simple production model. The Schaefer production model attempts to estimate r and K as its “leading parameters”. These can be easily transformed into the management variables MSY and F_{msy} as: $MSY = rK/4$, and $F_{msy}=r/2$ (see Hilborn and Magel, 1997). Simply solving the previous expressions for r and K results in $r=2F_{msy}$ and $K=2MSY/F_{msy}$. Given these simple analytical expressions, models can be fit to data by directly estimating MSY and F_{msy} and deriving r and K . Also, the implied priors for r and K can easily be calculated from the informative priors on MSY and F_{msy} . TINSS performs a similar variable transformation, but for a more complex age-structured model that might involve complex selectivity functions that may change over time.

The major problem that TINSS has fixed is that informative priors for B_0 and steepness are no longer required in data/information poor systems. Instead, the user must specify (in cases where there is a lack of contrast in the data) informative priors for MSY , or F_{msy} , or both. It might be much easier to use the sparse historical data to come up with an informative prior for MSY than it would be for B_0 (e.g. MacCall, 2009). In addition to priors, the other problem that TINSS addresses is the subjective weighting of age-composition data through the use of the conditional maximum likelihood estimates of the variance for age-composition data. To do so, the assumed statistical distribution for age-composition data are the multivariate-logistic distribution.

TINSS was developed specifically for the Pacific Hake fishery off the west coast of North America, and has not been used for the assessment of any other fish stock at this time.

Rapporteur’s summary of presentation

Not a generalized model, it was written specifically for Pacific hake, this was contract work for comparative model to Pacific hake SS model. Does not estimate B_0 or h ; MSY and F_{msy} are parameters of model and require priors. Uses Alec MacCalls DCAC to come up with informative priors for MSY and F_{msy} . Does not permit growth overfishing, can use multiple fleets but all of the selectivity must be the same, definition of recruitment is changed

What problem did this fix? No longer need to estimate B_0 and h . Using bicubic splines to reduce the number of selectivity parameters.

2.6.2 Summary of WKADSAM discussion

What comes out of model? Also get F , biomass, etc. Steepness and B_0 are derived values.

How does recruitment variability figure into year specific MSY calculation?

Q: Confused by outputs – are only $FMSY$ and MSY available?

A: No. Everything is available, but access to these is dependent upon developer

Q: RE: MSY . Are you assuming selectivity is fixed over time?

A: No. Corresponds to the initial year $FMSY$ & MSY . To know what they are today (i.e. terminal year) must compute B_0 and h for 1950, the run model forward and regenerate $FMSY$ and MSY .

Q: But MSY depends upon population size. How is recruitment variation connected to MSY given episodic recruitment?

A: Same as SS; I think. Can use SPR or MSY equilibrium. Doesn't use SPR40% as an MSY proxy.

Q: Does prior on MSY and FMSY differ over years?

A: Only if MSY units or metric changes over time does this matter. But this approach uses prior for initial year. Then, get marginal posterior for MSY, FMSY in terminal year.

Q: Could equally st prior for 2008. Would results be sensitive?

A: Yes, particularly if no info/contrast in the data. But this situation also occurs if setting priors upon then estimating B0 and h.

(Follow-up) Q: But B0 has only 1 interpretation – so illogical to consider any other year??

A: If you have no information or a 1-way trip you need informative priors to get it to work. Better to use informative priors as opposed to a single point-estimate. It is more honest representation, yet there are trade-offs.

Text on discussion:

A question was posed regarding the typical output from the TINSS software. The presenter noted that all quantities relevant are available in addition to MSY and FMSY, but software modifications may be required.

Most of the discussion centered on the parameterization of the model, particularly on how MSY and FMSY are the leading parameters. As such, priors are specified on these parameters as opposed to B0 and S-R steepness (h). It was argued that assessment practitioners are likely better able to construct reasonable priors for MSY and FMSY vs. B0 and h.

2.7 CSA

2.7.1 Description

In brief, Catch-Survey Analysis (CSA) is an assessment method that aims at estimating absolute stock abundance given a time-series of relative abundance indices, typically from research surveys, by filtering measurement error in the latter through a simple two-stage population dynamics model proposed by Collie and Sissenwine (1983 ; the acronym can also stand for Collie-Sissenwine Analysis).

CSA is one attempt to fill the gap between age-based methods, involving sequential population analyses for estimation, and biomass dynamic models (Conser, 1995). The former require relatively long time-series of catch-at-age data which (i) are generally expensive to process and more so as they need to be provided on a routine basis, (ii) must be reliable over the whole age range, and (iii) are simply unavailable for a number of species for which age determination is still an open question (e.g. crustaceans). Surplus production approaches may not provide reliable estimates when variations in stock abundance are more influenced by changes in recruitment than by response to fishing intensity, hence the addition of an explicit treatment for recruits in a two-stage model is a major improvement.

The starting point for CSA is that one is able to partition the survey (or cpue) time-series into two stages. The younger stage comprises the "recruits" which should be from a single year class. Hopefully, this group might be identified based on sparse age readings or by splitting length compositions at a distinct cut-off size. Animals of

all larger lengths or ages are accumulated into the second stage, called “fully recruited”, which is akin to a super plus-group. One must also provide an estimate of the total catch (fishery’s removal) in number in each year. The main advantage of CSA is that catch need not be subdivided by age; however, if the reported catches are inexact for any reason, CSA is not better off than VPA, say: CSA is not a fishery-independent method. In addition, one must specify the ratio of survey catchabilities (noted s) between recruits and fully-recruited which, in practice, is not estimable.

Properties of the method have been explored by simulation (Cadrin, 2000; Collie and Kruse, 1998; Mesnil 2003, 2005). The sensitivity of CSA results to errors in natural mortality, trends in survey catchability, misreporting of catches is qualitatively similar to that of VPA. For moderate levels of errors in the survey indices, CSA can reproduce trends in relative stock abundance reliably, but can go astray if the survey is very imprecise. Missing survey data do not have a dramatic effect, if infrequent of course. However, estimates of absolute abundance are very sensitive to the input catchability ratio s , for which no satisfactory method of determination exists at the moment; its value is not simply a function of gear selectivity but seems to ‘absorb’ other processes such as the relative variance between recruits and post-recruits. Nevertheless, comparisons with VPA indicate that both methods can produce very similar results. This shows that a full account of the age structure in the population is not essential for providing robust management advice.

The method is implemented in various pieces of software such as the NOAA toolbox or interactive R scripts. They are reputed to be very simple to use, given that the method has very few settings to fiddle with.

2.7.2 Summary of WKADSAM discussion

Benoit Mesnil presented the Catch-Survey Analysis (CSA) stock assessment method. The key discussion points raised following the presentation were:

- The presentation noted that CSA did not show a retrospective pattern even for the simulation case where survey catchability changed over time. The discussion suggested two possible reasons:
- (i) CSA uses only indices of abundance for recruits and the fully recruited animals and the total catch. The conflicts in the data and/or assumptions that cause retrospective patterns in fully age-structure models may not occur when data are aggregated at the level used in CSA.
- (ii) CSA is a forward-calculation model. Such models may be less likely to exhibit retrospective patterns than back-calculation models (e.g. VPA).
- CSA performed well in the simulation testing—the sole exception being its performance on the NRC Data Set 5. But even in this case, the CSA stock size trend matched the true trend fairly well.
- It was suggested that generally it may be better to retain the full age structure in the underlying population dynamics, then collapse the model predictions for use in the likelihood function, e.g. to get predicted indices for recruits and fully recruited. The full scope of this discussion was beyond the scope of CSA discussion and the WG decided to postpone this discussion for a later agenda item.
- It was noted that the Baltic Working Group has found that simple, two-stage models (similar to CSA) can give results comparable to fully age-structured models.

- The structure of CSA allows for both measurement and process errors. The various versions of the software implementing CSA have had difficulty estimating the respective variances. It may be fruitful to recast CSA as a random effects model with software implementation in ADMB-RE.

2.8 ADAPT-VPA

2.8.1 Description

The program Adapt VPA (virtual population analysis) is part of the NOAA Fisheries Toolbox (<http://nft.nefsc.noaa.gov>). This toolbox contains a wide range of stock assessment software packages and associated programs for stock assessment. Each program is a stand-alone Windows desktop application and combines a sophisticated graphical user interface (GUI) with an independent calculation engine. The GUI simplifies the data entry and provides tables and graphs of the output. Text files of output are produced from each program as well, with some programs directly creating rdat files for use in R. The programs range in complexity from simple index methods to complex age and length based stock assessments. The website also provides a population simulator that can interact with each of the stock assessment models for testing hypotheses regarding data uncertainty or model misspecification. Each program has a short description on the website and most programs come with user guides and technical reference manuals as well as at least one example input file. User support is provided via e-mail or phone. Many of the models in the toolbox are used in peer-reviewed stock assessments in the US and globally.

Adapt VPA has been the workhorse of the Northeast Fisheries Science Center for many years. It started as an APL program written by Ray Conser and has evolved to its current form as a Fortran 90 program with full graphical user interface with influences from Gavaris, Mohn, Powers, Restrepo, and Darby. As with all virtual population analysis programs, it performs best in situations of high fishing mortality rate and strong ageing programs due to the assumption of low error in the catch-at-age and the backward convergence property of cohort calculations.

Adapt VPA has a number of features that have been found useful in stock assessments in the Northeast US. The basic data are defined by a rectangular matrix of years by ages, with the oldest age optionally a plus group. This rectangular matrix must be filled for catch-at-age, three weights-at-age (catch, stock and spawning-stock biomass are entered separately and can be the same or different), natural mortality, and maturity. Survey or catch per unit of effort time-series are used to tune the VPA. These tuning indices can be entered in units of numbers or weight, all ages or specific age ranges, and relate to either the population at the start of the year or the mean population during the year. Weighting factors can be applied to each index observation. The program directly estimates a user defined number of stock abundances at age in the terminal year plus one and has multiple options for filling the ages which are not directly estimated. There are also multiple options for estimating the oldest true age and linking the plus group to the oldest true age. Catchability coefficients for each tuning index are calculated analytically. Optionally, catch multiplier parameters can be estimated for a set of years which allows for estimation of under or over reporting of catch.

The program can be run in five different modes. The simplest is the single run mode which produces the best point estimate for the parameters. Bootstrapping of the tuning indices allows non-parametric estimation of the uncertainty associated with the parameter estimates and terminal year metrics such as spawning-stock biomass, total

biomass, fishing mortality rate, and recruitment. Due to the backwards convergence property of VPA, the uncertainty estimates become vanishingly small for earlier years in the assessment time-series. The program has a built-in retrospective run mode which removes years of data one at a time from the most recent backwards and re-runs the model for each “peel.” A combined bootstrap and retrospective run mode bootstraps each retrospective “peel” to examine the significance of retrospective patterns. The sensitivity analysis run mode varies the natural mortality rate for all or a portion of the year by age matrix systematically.

The Adapt VPA graphic user interface provides a number of standard plots and allows the user to examine the full output file. The standard plots include time-series of spawning stock or total biomass, population abundance-at-age, fishing mortality rate, partial recruitment, observed and predicted tuning indices, and residuals for the indices. Bootstrapping, retrospective, or sensitivity run modes create additional plots and output. The results can also be exported to an rdat file for use in the R software package.

The advantages of Adapt VPA include the graphic user interface to ease data entry and examination of output, fast run speeds, ease of use, simple formulations make it easy to understand and explain to review panels and managers, and the many built-in options such as the retrospective analysis and bootstrapping. However, as with all virtual population analyses, the program requires age only data, cannot internally estimate reference points, and cannot estimate the uncertainty of the early years in the time-series due to the backwards convergence properties of VPA. The fast run speeds and ease of use have allowed the program to be used to examine a number of stock assessment issues. For example, the bootstrap and retrospective run mode allows for an objective determination of whether or not an observed retrospective pattern is strong or not based on the amount of overlap of the bootstrapped distributions. The program was also used in a moving window analysis to detect the time of an intervention leading to retrospective patterns for a number of groundfish stocks. The program has been incorporated in a management strategy evaluation software package and used to evaluate the effectiveness of splitting survey time-series as an approach to address retrospective patterns caused by different sources.

Rapporteur's summary of presentation

Introduction to NFT

This is a collection of software programs from really simple to highly complex models. The models are used globally. Available from the website are stand alone executables as well as windows GUI to help facilitate users and to allow for some data checking. Graphics and export to R are also provided with many of the programs.

Website: <http://nft.nefsc.noaa.gov/>

Basically the user sees the GUI and the background is the data entry, plot files etc. There are also reference manuals available for each program.

Model comparison list

Also set up a list of different features by models available to show which features are incorporated within a particular model. This is somewhat a comparison of models to Stock Synthesis because this has all the features listed. Within each model there is a write up with information about the model, references, and download capabilities. Also some user support for many of the models and the NFT as well. Each of the programs has the same look and feel for easy comparison between each.

Chris next presented two of the programs in NFT.

ADAPT-VPA

This model performs best with high F and strong ageing program. It is considered the workhorse of NEFSC and is quite easy to use. Technicians running these models can use a point and click interface because may not be as highly trained in assessment. The estimation procedure is the Levenburg-Marquardt (IMSL). Chris stated that robustness is “high” because it’s actually difficult to break the VPA. He considers robustness in the sense of ability to run and produce sensible results. This is definitely a data-rich method. It requires tuning indices and is very difficult to use in data-poor situations. Designed for assuming error in catch-at-age is low relative to error in tuning indices. Uncertainty is derived through the Hessian. Bootstrap routine also available. Length must be converted to age.

Pro/Con: very fast, easy to use, easy to explain. Easy to check different scenarios and change data. The built in retrospective analysis is a nice additional diagnostic. However, this is a purely age-based model, should not use this if do not have good ages or no ages. There are relatively few options and this would be considered a “stiff” model. Cannot internally estimate a reference point, have to do this afterwards.

Problems Solved:

- 1) Retro bootstrap: when trying to determine if the retrospective analysis is strong and need to do something about this.
- 2) Moving window analysis: when trying to decide when a large change has occurred.
- 3) Split surveys: to reduce retrospective analysis if a change has occurred.
- 4) Missing catch: also to reduce retro pattern.
- 5) Weighting indices by year: recently added to deal with specific issues in data

Chris then showed examples of each of these problems solved.

Retro bootstrap: Good to evaluate the strength of the retrospective pattern. Suggested that no overlap of time-series means very strong retrospective pattern and likely need to consider what is causing this.

Moving window: Presented normalized values of a parameter (in this case q) and simulated a strong intervention in a particular year. Moving window process was able to detect this change. Once an issue is identified then can make a change such as splitting the surveys. Then rerun the retrospective analysis and see if the pattern reduces. The fix may help solve problems in catch advice even if the fix is for the wrong mechanism such as splitting surveys and changing q , even if the problem was with M .

Splitting Surveys: Question was posed on how to make a decision to split surveys when a retrospective pattern was detected. There are a few things that can be done to deal with the retrospective pattern (e.g. split survey, change catch). There may be no real evidence for these particular issues, because there could be a change in M , but this provides a way to deal with the retrospective pattern.

An example was presented where he had SSB and F and did a number of sensitivity runs within the range of uncertainty. When he did not split the series, perception of stock size was high and F was low. This created a retrospective pattern because every year realized that the stock was not that high and F was not that low. When split the

surveys, the F was much higher, so he could explain the higher catch. With this model you do need to make a choice as to which fix you should do.

2.8.2 Summary of WKADSAM discussion

Doug: This seems like a tactical ad-hoc fix, and not an explanation of what happened. Loose precision when throw data away and haven't actually figured out what the real problem is. These fixes might be fine in the short-term decision-making, but not that great for long-term and for deciding what the reference points are.

Noel: Seems a bit of an issue for the short-term issue as well. Are you redoing the VPA every year?

Chris: No, this is a two year projection, so started out being too conservative then balanced out.

Problem Indices: Demonstrate what can happen to indices when using simple models without catch. Chris showed four different surveys. One survey had individual tows that caused the values to quadruple in size. With catch-at-age, this is easy to deal with, but with VPA was not previously easy to deal with. Now can downweight the indices to deal with problems such as these.

Doug: This is something of a discussion for later on how to deal with surveys showing high increases in catch. Should someone deal with these years or just ignore them. Reality is that will get survey indices like this when have species that are highly aggregated. Unbiased estimate is not necessarily a very good estimate when have a highly skewed distribution. Certainly have to do something, but can we just treat them as IID or throw away as outliers.

Chris: This is something to think about because when we simulate data that behave well then all the models will perform well, but that's not reality.

2.9 ASAP

2.9.1 Description

The program ASAP (age structured assessment program) is part of the NOAA Fisheries Toolbox (<http://nft.nefsc.noaa.gov>). It is a relatively simple statistical catch-at-age model with a graphical user interface. The program was originally developed for use in ICCAT in the late 1990s, was modified for use in the NOAA Fisheries Toolbox in the early 2000s, and was updated in the mid 2000s for added flexibility. ASAP was one of the first AD Model Builder stock assessment programs to fully utilize the flexibility of ragged arrays (J. Ianelli, pers. comm.).

The following description assumes use of the GUI with sections on what is found in each tab and how they relate to each other. The General Data tab contains the basic information necessary to dimension the problem, the number of years, ages, fleets, available surveys, and selectivity blocks. A number of new features are contained within the General Data tab. One is the ability to enter all the available indices for the stock assessment then select which ones to use in a particular run in the Index Specification tab. This facilitates running the model with and without given indices rapidly. Another new feature is the use of selectivity blocks which each contain their own parameters and initial guesses. For example, selectivity can now be fixed in the earliest and most recent years, but estimated in the remaining years, by fleet. The inclusion of discards as observations is now treated through a check box. This means that when discards are combined with landings to produce total catch, as commonly done in many assessments, the user does not need to enter two large matrices of ze-

ros as well as a number of other parameters. Full calculation of all likelihood terms in the objective function can be requested by checking the Use Likelihood Constants box. If this box is not checked, then the constants in the lognormal and multinomial error distributions will be ignored. Typically, this box should be checked. Another new feature of ASAP is the ability to skip projections by not checking the Perform Projection Calculations box. The final new feature of ASAP shown on the General Data tab is the optional calculation of Monte Carlo Markov Chain (MCMC) calculations to estimate uncertainty in the model solution. While MCMC is always available to any ADMB executable, it could not be accessed through the original ASAP GUI.

The Biology tab contains three sub-tabs: Natural Mortality, Maturity, and Fecundity. The first two are simply year by age matrices. The Fecundity sub-tab contains the same option as the original ASAP regarding the use of spawning-stock biomass (select Multiply Maturity by Weight at Age) or fecundity (select Use Maturity Values Directly and enter eggs per adult in the Maturity sub-tab). A new feature in the Fecundity sub-tab is the definition of when during the year spawning occurs. The estimated population will be reduced by that proportion of total mortality for each age and year prior to calculating the spawning-stock biomass (or number of eggs if Use Maturity Values Directly option selected). Note that ASAP does not distinguish males and females, so if only female spawning-stock biomass is desired then the maximum proportion mature should be the sex ratio at age.

The Weights at Age tab also contains three sub-tabs: Catch Weights, Spawning Stock Weights, and JAN-1 Stock Weights. Each is a year by age matrix and they differ only by where they are applied. The Catch Weights are used to calculate the total catch and discards, yield-per-recruit, and projected yields. The Spawning Stock Weights are used to calculate the spawning-stock biomass (SSB), if selected in the Fecundity sub-tab of the Biology tab. The JAN-1 Stock Weights are used in the biomass average for Freport and for indices with biomass units. The three matrices can contain identical values or different values depending on the data available.

For each fleet, a number of specifications are entered in top portion of the Fleets tab: a name for the fleet, selectivity starting and ending ages, release mortality, and a fleet directed flag. The starting and ending ages for selectivity determine which ages are used when comparing the observed and predicted proportions catch and discarded at age. These ages also determine the ages over which the total catch and discards in weight are computed. Thus, the starting and ending ages will commonly be 1 and the maximum age, with the exceptions being cases where specific fleets never catch young or old fish. The Selectivity Block Assignment portion of the Fleets tab contains sequential integers which define the years and fleets where selectivity is the same. The blocks will almost always be sequential years within a fleet, but two fleets could be assumed to have the same selectivity for a number of years. The final information entered on the Fleets tab relates to the new Average F feature of ASAP whereby a starting and ending age define a range of ages for which to report the total fishing mortality rate. This average F approach is similar to the one in the NOAA Fisheries Toolbox Adapt VPA program and facilitates comparison of fishing mortality rates over years when selectivity patterns or fishing intensity by different fleets change.

Each selectivity block is entered one at a time controlled by a drop-down box in this tab. For each block, there are three selectivity options: By Age, Single Logistic, and Double Logistic. Depending on the choice made in the Selectivity Option drop-down box, the Selectivity Specification area will change to show only the type of data needed. Whichever selectivity option is chosen, there are four columns of information

to be entered: Initial Guess, Phase, Lambda, and Coeff. of Var. Note that any combination of estimated and not estimated parameters is allowed.

Catch information is entered by fleet according to the drop-down box at the top of the Catch tab. For each fleet, the Catch at Age contains the information that will be used to create the observed proportions at age. The Total Weight for each year will often be in units of metric tons. Note that the units for the estimated numbers-at-age in the population will be determined by the units used for the catch weight matrix (on the Weights at Age tab) and the units used for total weight here.

The Discards tab is arranged exactly like the Catch tab. If discards are entered on this tab, then the catch tab should have only landings entered, do not enter catches twice. Only dead discards should be included in this tab, not discards that survive.

The Release tab is also entered fleet by fleet, similar to the Catch and Discards tabs. However, only the proportion at age and year released is entered, there is not a second entry box. These proportions will determine the fate of catches. The number released will be multiplied by the release mortality to produce the number of predicted discards to compare to those entered in the Discards tab.

Seven pieces of information for each index are entered on the Index Specification tab. The Index name is self-explanatory. The Units are 1 for biomass and 2 for numbers. The start of a month is used to determine the timing of the index during the year. The Selectivity Link to Fleet allows the selectivity for an index to depend on the selectivity of one of the fleets, for example due to the index being a catch per unit of effort from that fleet. The Selectivity Start Age and End Age determine the age range to which this index applies. Indices can come from all ages or from just a single age, as either a recruitment index or as an age-based index similar to many VPA applications. If the latter is the case, then the start and end age will be the same number. Finally, a one is entered in Use Index in Estimate if the index is to be included in fitting the model, a zero is entered to ignore the index.

The Index Selectivity tab is arranged similarly to the Selectivity Blocks tab, whereby each index is entered one at a time with the three options for selectivity. When Selectivity Start Age is the same as Selectivity End Age in the Index Specification tab, the GUI automatically sets the Selectivity Option for that index to By Age and fills the initial guesses with zeroes for all ages except the Start (=End) age which it fills with a one, sets the phases for all ages to negative one, all lambdas to zero and CV to one. A note appears under the selectivity option drop down box when an index is linked to a fleet. In this case, the selectivity option is irrelevant, but whichever option is selected the user should ensure that all phases are set to negative values.

The Index Data tab also treats information for each index separately using a drop-down box at the top of the tab. When no selectivity parameters are estimated for an index, the only information that needs to be entered are the annual values and associated coefficients of variation which determine how closely the model will try to match the index values. When one or more selectivity parameter is estimated for an index, then proportions at age along with an effective sample size must also be entered for each year. The GUI will automatically hide these columns if they are not required. If the GUI is showing these columns, then the information in them will be used in calculating the objective function. If index selectivity parameters are estimated, but all effective sample sizes for that index are set to zero, there will not be any information to estimate the parameters and the program may crash or be unable to invert the Hessian.

Phases are used in ADMB to help the model get into the correct solution space using a limited number of parameters then slowly adding parameters. Negative values mean that a parameter is not estimated, but rather fixed at its initial guess. Positive integers are used to determine when in the estimation process a given parameter becomes estimable. All parameters from previous phases are still estimated as the program moves to the next phase. Generally, it is recommended to set the scaling parameters to early phases and deviation parameters to later phases, when estimated in the Phases tab.

The three Lambda tabs all deal with how much emphasis to place on different portions of the objective function. A lambda of zero means that portion does not add at all to the objective function. In most cases, it is considered best practice to set all non-zero lambdas to one for base case runs and use a range of lambda values for a specific component to conduct sensitivity analyses. The total catch and discards in weight information is entered in the Lambdas-1 tab. The first piece of information necessary is the CV (coefficient of variation). Note that large CV values for large components of the total catch can cause the model to become unstable. The second matrix is the input effective sample size which is used in calculating the contribution to the objective function from the proportions at age for catch or discards. Entering a zero for a fleet and year combination means those proportions at age will be ignored. General rules of thumb for non-zero effective sample sizes are 50–200, although particularly bad or good sampling can lower or increase this range, respectively.

The Lambdas-2 tab contains the recruitment information. The lambda for recruitment deviations refers to the deviations between the recruitment predicted from the Beverton and Holt stock recruitment relationship and the estimated recruitment in that given year. Setting the lambda to zero means the model will be unconstrained with respect to estimating recruitment, provided the phase for recruitment deviations in the Phases tab is positive. This is not recommended, as solutions with one extremely large cohort often result. Instead, set the lambda to one then use a large CV if the desire is to have minimal constraint on recruitment estimates. The annual recruitment CVs can also be used to cause the model to estimate recruitment directly from the B-H curve, for example in early years when no age information is available, by setting the CV to a small value, such as 0.001.

There are three separate sections on the Lambdas-3 tab, which depend on whether the components are fleet based, index based, or neither. For each component, a lambda is entered and an associated coefficient of variation.

Values to start the program are entered in the Initial Guesses tab.

When projections are conducted, the number of years to project is determined by the top box in the Projection tab. For each year, recruitment can either be directly from the stock recruitment relationship or else be a user supplied value. The rule for each year is one of five options: match an input yield, fish at a specified $F\%SPR$, fish at F_{msy} , fish at the current F , or fish at an input F . The final column is a multiplier for the bycatch fisheries to control whether they remain the same (1.0), increase (value > 1.0) or decrease (value < 1.0) in the projection years. These are simple, deterministic projections. For stochastic projections, the NOAA Fisheries Toolbox program AgePro is recommended with output from ASAP provided through the MCMC option.

Monte Carlo Markov Chain is a common approach to estimate uncertainty in models. A simple MCMC call is implemented in the MCMC tab requiring three values to define its operation: the number of MCMC iterations to output, the thinning factor for the MCMC algorithm, and a random number seed. The GUI will compute the total

number of MCMC calculations as the product of the first two values above then save only one value for every of the second value.

The ASAP program can be run in three distinct ways: the basic run, with MCMC, and with a retrospective analysis. These three options are available in the Run menu, and each has a corresponding ability to scan back the results of a previous run, so that a long run does not need to be repeated. The basic run conducted the standard ADMB minimization of the objective function and outputs the report file along with a number of helpful graphics in the GUI. Running ASAP with MCMC produces Monte Carlo Markov Chain estimates of uncertainty for the F and SSB time-series as well as the MSY related reference points and a file for input to the AgePro projection program. Running ASAP with a retrospective analysis produces a number of input and output files, the results of each can be viewed individually by opening them in the GUI, and summarizes how the estimates of F, SSB, and N at age change as years of data are removed from the model. This retrospective analysis can help identify problems in the data and assumptions that cause successive estimates to be biased as years of data are added or removed.

The output is arranged in separate tabs in the GUI, with the full report file available for examination as well. The Model Fit Summary tab contains a summary report of the different components comprising the total objective function. A number of common fishing mortality based reference points are provided: $F_{0.1}$, F_{max} , $F_{30\%SPR}$, $F_{40\%SPR}$, and F_{msy} . All five reference points are calculated assuming the selectivity pattern, weights-at-age, natural mortality rate, and relative fishing intensity among fleets in the terminal year of the assessment. Parameters of the estimated or assumed stock recruitment relationship are presented along with a graph. A plot compares the input and estimated effective sample size (ESS) by year for each fleet for both catch and discards. When input ESS is zero, the estimated ESS values have no real meaning and can be ignored. A final assessment should have some similarity between the input and estimated ESS, although the values for each year are not expected to be exactly the same.

Time series plots for each fleet comparing the observed and predicted total catch or discards in weight or else the standardized residuals of these values. Standardized residuals are computed in log space based on the user input CV for each catch or discard value. Similar plots are produced for the fitted indices. Two types of selectivities are shown: fleet specific and index specific. The fleet specific selectivity plots are three-dimensional showing changes over both age and year (year specific values depend on the selectivity blocks defined on input). The index specific selectivity plots show only selectivity at age under the assumption that all years have the same selectivity. This assumption is broken when index selectivity is linked to a fleet selectivity, in which case only the first year selectivity is shown in the plot and the user must examine the correct fleet selectivity associated with that index.

Directed, but not discard, fishing mortality-at-age and year by fleet and total are plotted in separate tabs. This allows comparison among fleets of the fishing mortality associated with the landings both in trend and magnitude over ages and years. The average fishing mortality rate is computed three ways: an unweighted average over the age range, weighted by associated population abundance in numbers (N-weighted), or weighted by associated population abundance in biomass (B-weighted). When selectivity is the same for the age range selected, these three average F values will be the same. However, when the selectivity differs by age within the age range selected, the values can be quite different.

The population numbers-at-age and year matrix can be viewed two ways in this tab: at age for a specific year or overall years for a specific age. The check box to set the y-axis to maximum in all years can be helpful for detecting large cohorts passing through the matrix. Index specific catchability coefficients are plotted over time for all years where positive index values were entered. This plot is typically only used when catchability deviations are allowed so that catchability changes over time, otherwise it will simply be a flat line showing which years had information for each index.

Proportions at age for catch or discards by fleet and year are plotted as observed and predicted values in this tab. Note that the header of the graph provides the input effective sample size for that fleet and year. When input ESS is zero, the comparison between observed and predicted proportions at age is not meaningful because it has no impact on the objective function. At the bottom of the Year pull-down menu, is a total which computes an average proportion at age for both observed and predicted values from only years for that fleet when input ESS is greater than zero. This is a simple overall comparison that has no meaning in terms of the objective function, but provides a quick summary of how well the fleet is fit over all years included in the fitting process.

The Projections tab provides in graphical form all the output related to ASAP projections. There are seven plots with age as the x-axis and separate lines for each projection year: population numbers, directed F, bycatch F, non-directed F, catch (in numbers), discards (in numbers), and yield (in weight). These plots are followed in the drop down list by four plots with year as the x-axis summarizing annual totals: total yield (in weight), total discards (in weight), spawning-stock biomass, and the ratio of spawning-stock biomass to SSB_{msy}. Note that the projections in ASAP are relatively simple and do not include the uncertainty of current abundance-at-age or future recruitment. Use of AgePro (available in the NOAA Fisheries Toolbox) is recommended for more advanced stochastic projections.

There are four plots presented in the Time Series tab: total stock numbers, spawning-stock biomass, estimated recruitment (denoted observed recruitment), and recruitment expected from the stock recruitment relationship (denoted predicted recruitment). The spawning-stock biomass and observed recruitment plots also show the estimated uncertainty based on the .std file of plus and minus one standard deviation.

One way to get started quickly with ASAP is to import a virtual population analysis input file. This option is available under the File pull-down menu and assumes the latest NOAA Fisheries Toolbox Adapt VPA software was used to create the VPA input file. The import will create an ASAP input file that should run. However, the conversion program is relatively simple and users should carefully examine all inputs to ensure they are correct, especially lambda and CV values.

A file called "asap2input.log" is created each time the program is run. It reproduces the input as the input file is read and can be used to find problems in the input file itself. If the "asap2input.log" contains different information than your input file, it means an error was encountered while reading the input file. This should not happen when the GUI is used, but can occur if the input file is edited by hand outside the GUI.

As with all statistical catch-at-age models, there are many possible ways that the estimation process can produce unreliable results. If the model is not able to estimate the Hessian (.std file not produced), this is an indication that the minimization of the objective function encountered a problem and results are suspect. This can happen

when parameters are estimated but no data are provided to inform the estimation, for example, if index selectivity is estimated but all the input effective sample sizes are zero for that index. Careful review of input settings should allow detection of this problem. Alternatively, changing the number of parameters estimated (by setting phases negative) can allow determination of which parameters are causing estimation problems. Using lambdas with loose CVs can sometimes allow estimation of parameters in cases when unconstrained estimation does not work.

Once a model does run to completion and the Hessian is inverted, there is still a question regarding the suitability of the model. Standard diagnostics, such as runs of residuals, should of course be examined. If no standard diagnostics show a problem, then next step is to determine if the model inputs and outputs are compatible. A fully compatible model will have the input CVs exactly matched by the uncertainty in the output fits. This can be seen in both the effective sample sizes and the standardized residuals of fits. In reality, neither will ever match exactly due to random chance, but the overall level should be compatible. When comparing the input and output effective sample sizes, the pattern and overall magnitude should be similar. However, there will be occasional cases when the output effective sample size becomes very large relative to all other years. This is just an artifact of the calculation of output effective sample size and can be ignored. The standardized residuals should all follow a normal distribution with mean zero and standard deviation one, meaning approximately 95% of the observations should fall within plus or minus 1.96. Another way of summarizing the standardized residuals is to compute the root mean square error (rmse) for the set of residuals. An infinite sample drawn from a $N(0,1)$ distribution has rmse equal to one. However, when sample sizes are limited, as they often are in fisheries assessments, rmse values drawn from a true $N(0,1)$ distribution can have a relatively wide range. This means that the goal of “tuning” an assessment should necessarily be to get the rmse to exactly one, but rather to have the rmse fall within the confidence bounds associated with a $N(0,1)$ distribution for that sample size. These sample sizes and rmse values are given in the ASAP report file just before the projection results to help guide the “tuning” of the assessment. For example, if rmse is too low, the CV can be reduced, while if the rmse is too high, the CV can be increased. Alternatively, the model can be restructured such that the input and output CVs coincide. The former ad hoc approach to tuning is used in many assessments, but has also been criticized by some for its irreproducibility. The latter approach is preferred, but may not be achievable if the input CV's are too tight relative to all possible models due to the lack of process error in their calculation. This topic remains an active area of research in fisheries stock assessment.

Once a set of parameters have been found which appear to be reasonable, it is recommended that a series of sensitivity runs be conducted to examine how sensitive the results are to some of the input choices. One way to do this is to conduct a likelihood profile over a range of possible values for a parameter. Either of the stock-recruitment parameters are often useful for this exercise. Fix the parameter at a range of values both above and below the estimated value and compare the objective function components. This exercise will often identify which pieces of information are in agreement, the objective function components follow a similar trend, and which pieces of information are in conflict, the objective function components are minimized at different parameter values and follow different trends. This is often useful to help define specific sensitivity analyses which either remove or significantly downweight specific pieces of information to show how influential they are on the

final result. Stock assessments with highly conflicting basic data will require either choosing one set over the other or else provision of advice based on both possibilities.

Rapporteur's summary of presentation

Chris next presented this statistical catch-at-age model, and they considered this a gateway model for getting people from going backward to going forward.

Lineage of ASAP: This model is one of the first ADMB SCAA models to be used and was developed originally for ICCAT in late 1990s. More flexibility was added in mid 2000s. The increase in computing power has allowed for these models to have more flexibility.

Features: Ease of use is moderate as well as robustness because there are ways to not make it work. Easier to handle uncertainty in catch-at-age, because explicitly dealt with in model. Model is more appropriate for data poor situations, but is still strictly an age based model. Need to have conversion to age data. As with ADAPT, one area, one sex model.

Pros/Cons: relatively fast, easy to use, flexible, calculates internal reference points and built in retrospective analysis. This is done automatically in the GUI with a peel and can look at all the runs individually. However, purely age based, must convert lengths to ages, limited in parameter estimation.

Problems solved: helps with moving VPA to SCAA, also talks to standard AgePro software with which managers are accustomed to using. Managers can't really tell the difference in output between VPA and SCAA. Also has RMSE as indicator of fish and effective sample size. SCAA does not really make retrospective patterns go away. An additional feature is that user can import a VPA file to populate the ASAP model which is a nice time saver.

Chris then showed an example using pollock, where selectivity on 10+ group was changed from 0.1 to 1. In this way he changed one parameter and explored results of this decision. Pollock working group example showed a bundle of different parameters by model runs and comparison of values across models. Because the model runs very fast, can explore lots of space in terms of potential settings.

AgePro was then used to show the probability of stock being overfished into the future. This allowed for gaining some ground on moving managers toward supporting the use of SCAA models. Can also use RMSE as indicator of fit (should be near one), as well as effective sample size as a diagnostic.

Important to note that retrospective patterns can still be an issue regardless of whether using VPA or SCAA, doesn't magically make the retro pattern go away.

"Two quite opposite qualities equally bias our minds, habits and novelty."

2.9.2 Summary of WKADSAM discussion

Discussion Session

Doug: How reliable is the bootstrap estimate from these methods? Problem being lots of correlations among parameters. Had a situation where comparable models with extra parameter were getting worse fits. How many people have actually checked that their bootstrap values were reliable? Can get situation where bootstrap sticks and get a very narrow exploration of space. Comment is not model specific, just raising a general flag to be careful about bootstrapping when have multi-parameter estimation and quite a lot of covariance. Wondered if others had had similar experiences.

Ray: good point for automated procedures where don't have ability to explore what is going on. But with VPA, parameters are rarely correlated only dealing with survivors in terminal year and catchability.

Doug: rather than exploring what might be a very flat surface, the bootstrap might stick and not move from starting values.

Chris: definitely an issue when have lots of parameters, but with VPA not usually an issue and start bootstrap from wild values, doesn't really have problem getting back.

Noel: Can you comment on the Woods Hole experience with this?

Chris: they were finding out that there was high ageing error and that fishing mortality had gone down quite a bit, so wanted to move away from VPA. People are also starting to explore SS, because that allows for not having comfort with age data and can look into using lengths not strictly ages. This process is something of a progression.

2.10 SURBA

2.10.1 The development of SURBA

2.10.1.1 RCRV1A and earlier separable models

The basis of SURBA is a simple survey-based separable model of mortality. This was first applied to European research-vessel survey data by Robin Cook (1997, 2004), but the underlying model has a long history in catch-based fisheries stock assessment (e.g. Fournier and Archibald 1982, Pope and Shepherd 1982, Deriso *et al.*, 1985, Gudmundsson 1986, Johnson and Quinn II 1987, Patterson and Melvin 1996): see Quinn II and Deriso (1999) for a summary. Cook (1997) took existing separable methods and developed a way to apply them to survey data alone; all existing approaches had been based on catch data, or a combination of catch and survey data. Cook called his implementation RCRV1A, standing for (I think) Robin Cook Research Vessel 1A, and used it to estimate population trends for a number of European stocks. However, the program was not made generally available, and its use didn't extend much past the original paper and a follow-up (Cook 2004).

2.10.1.2 SURBA 1.0

During preparation for the ICES Northern Shelf assessment working group in Copenhagen in 2002 (ICES-WGNSDS 2002), it became clear that population signals arising from catch and survey sources were quite different, particularly for gadoid stocks in the West of Scotland and Irish Sea (ICES Divisions VIa and VIIa). This was evident from trends in survey residuals from catch-based assessments, but it was not obvious what the population trends would be if based on the surveys alone. To address this, we picked up where Robin had left off and put together an updated version of his model, restructured in Fortran-90 (although still without a Windows interface, and using only one survey at a time). This was SURBA version 1.0, which became SURBA version 1.03 during the meeting (Appendix 1, ICES-WGNSDS 2002). The implementation was rough and ready, but the results appeared to concur with the group's prior perception and we decided it would be worthwhile to take the model further. Version 1.0 subsequently became the basis of a paper on survey-based analysis for haddock in Division VIa (Beare *et al.*, 2002, Beare *et al.*, 2005).

2.10.1.3 Subsequent versions and use in ICES, and related models

In the years since 2002, I and others have used SURBA many times and in many different contexts. On a regular basis, someone will encounter a problem with the program, or apply it in a situation which I had not envisaged, and feedback from these occasions has helped to improve the code a great deal. An early development was the inclusion of many different surveys, rather than just one. The use of the program in assessment working groups increased rapidly after I developed the Windows front-end – which just goes to show that we fisheries scientists will use any model as long as it is easy to operate and produces nice pictures! Some other developments and applications included:

- During work for the ICES Methods Working Group, I investigated the possibility of using survey catchabilities estimated via a catch-at-age analysis to raise the relative stock estimates produced by SURBA to a pseudo-absolute level (Needle 2004a). The idea is to fit a catch-based VPA to catch and survey data from a period when catch data are thought to be reliable. The catchabilities determined in the fit can then be fed into SURBA, and the population estimates derived will be on the same scale as the historical population estimates from the VPA. The approach never really caught on, but much the same kind of thing is done now for cod in the North Sea using B-ADAPT (ICES 2008a), and for cod and haddock in the West of Scotland (ICES 2008b).
- SURBA has been subjected to testing using simulated data on a number of occasions (ICES, 2004a, ICES, 2006b, Mesnil *et al.*, 2008, Needle, 2004c, Needle, 2004b).
- SURBA formed part of the EU FISBOAT project, which ran for three years and included a number of European institutional partners. FISBOAT looked at the utility of providing management advice based on fishery-independent assessment models, of which SURBA was one example (Cotter *et al.* 2007b, Cotter *et al.* 2007c, Mesnil *et al.* 2008, Pomarede *et al.* 2006).
- Concurrently, SURBA has also formed part of the EU EFIMAS project (EFIMAS 2007). This project developed operational tools for evaluating management strategies, using a bespoke R library (FLR Team 2006). Laurie Kell (Cefas, Lowestoft) and I wrote an R function to enable SURBA to be run as part of this library.

SURBA is now used for two main purposes in ICES assessment working groups: to supplement existing catch-based VPA-type analyses, and to provide the basis for advice for a number of stocks for which catch data are not thought to be reliable.

The widely available implementation of SURBA (version 3.0) is coded in Fortran-90, using NAG library routines for parameter estimation and featuring a Windows GUI. There is also a module in FLR (FLSURBA) which uses the same estimation code as SURBA 3.0, but is no longer actively supported. Experimental versions are also being developed in R (SURBAR), SAS and ADMB (SURBA+).

2.10.2 The SURBA method

2.10.2.1 Basis

The separable model used in SURBA assumes that total mortality $Z_{a,y}$ in age a and year y can be expressed as

$$Z_{a,y} = s_a \times f_y$$

where s_a and f_y are respectively the age and year effects of mortality. Note that this differs from the usual assumption in that total mortality Z is the quantity of interest, rather than fishing mortality F . Then, given $Z_{a,y}$, abundance $N_{a,y}$ can be derived as

$$N_{a,y} = N_{a_0, y_{c_0}} \exp\left(-\sum_{m=a_0}^{a-1} \sum_{n=y_0}^{y-1} Z_{m,n}\right) = r_{y_{c_0}} \exp\left(-\sum_{m=a_0}^{a-1} \sum_{n=y_0}^{y-1} (s_m \times f_n)\right)$$

where a_0 and $y_{c_0} = y - a - a_0$ are respectively the age and year in which the fish measured as $N_{a,y}$ first recruit to the observed population (y_0 is the first year of the dataset). Thus the abundance at each age and year of a cohort is given by the recruiting abundance (denoted by $r_{y_{c_0}} = N_{a_0, y_{c_0}}$) of the relevant cohort modified by the cumulative effect of mortality during its lifetime.

2.10.2.2 Model parameters and procedures

The parameters to be estimated when fitting the model are $\Theta = [\mathbf{s}, \mathbf{f}, \mathbf{r}]$, where:

$$\begin{aligned} \text{Age effects:} \quad \mathbf{s} &= [s_{a_0}, s_{a_0+1}, \dots, s_A] \\ \text{Year effects:} \quad \mathbf{f} &= [f_{y_0}, f_{y_0+1}, \dots, f_Y] \\ \text{Cohort effects:} \quad \mathbf{r} &= [r_{y_0-A+a_0}, r_{y_0-A+a_0+1}, \dots, r_{Y-a_0}] \end{aligned}$$

Here A is the oldest age and Y is the final year. As presented, the SURBA model is indeterminate and has no unique solution. For this reason, elements of \mathbf{s} and \mathbf{f} must be fixed beforehand. Given a reference age a_r chosen by the user, the fixed values are:

$$\begin{aligned} s_{a_r} &= 1.0 \\ s_A &= s_{A-1} \\ f_Y &= \frac{1}{3} \sum_{y=Y-3}^{Y-1} f_y \end{aligned}$$

SURBA uses principally age-structured research-vessel survey indices as input data: biomass indices can also be used, but at least one age-structured index must be available. Data on stock weights-at-age and maturity are also required if relative SSB estimates are to be produced, while natural mortality estimates will enable estimation of F rather than Z . The user must provide survey catchabilities (often assumed to be constant due to lack of information) and SSQ weightings, along with the weight for the smoothing term on year effects (see below).

The fitting procedure is as follows. Indices are mean-standardized, so that the mean across all observations for a given index is 1.0. Age-structured survey indices are back-shifted to the start of the year, to enable comparison with abundance estimates for 1 January, while biomass indices are shifted to spawning time to enable comparison with SSB. Fitted abundance (and hence survey-index values, given catchability assumptions) are generated using the equations given above, and the sum-of-squares for age-structured indices is given by:

$$SSQ_I = \sum_{i=1}^{NI} \sum_{a=a_0}^A \sum_{y=y_0}^Y \omega_{a,y,i} (\ln I_{a,y,i} - \ln \hat{I}_{a,y,i})^2$$

where $\omega_{a,y,i}$ are user-defined weights. There is also an equivalent SSQ_I term for biomass indices, and the smoother in year effects:

$$SSQ_{\lambda} = \lambda \sum_{y=y_0}^{Y-2} (\hat{f}_y - \hat{f}_{y+1})^2$$

in which λ is the user-specified smoother weight. The final SSQ to be minimized is then the sum of these three terms. Minimisation is performed by least-squares regression assuming lognormal errors. Confidence limits on management outputs are estimated by the delta method (SURBA 3.0, for Z and recruitment only) or by parameter resampling from a multivariate Normal distribution defined by the inverse Hessian (SURBAR).

The main problems with SURBA are that it produces relative abundance estimates only (although this is the case for all survey-based approaches), it requires assumptions for the final year and oldest age, and it assumes separability in mortality. The mortality uncertainty estimates can also be very wide in SURBA 3.0. Finally, the other versions mentioned above are mostly in an early stage of development and have not been properly tested.

Figure 2.10.1 gives an example of output from the SURBAR implementation.

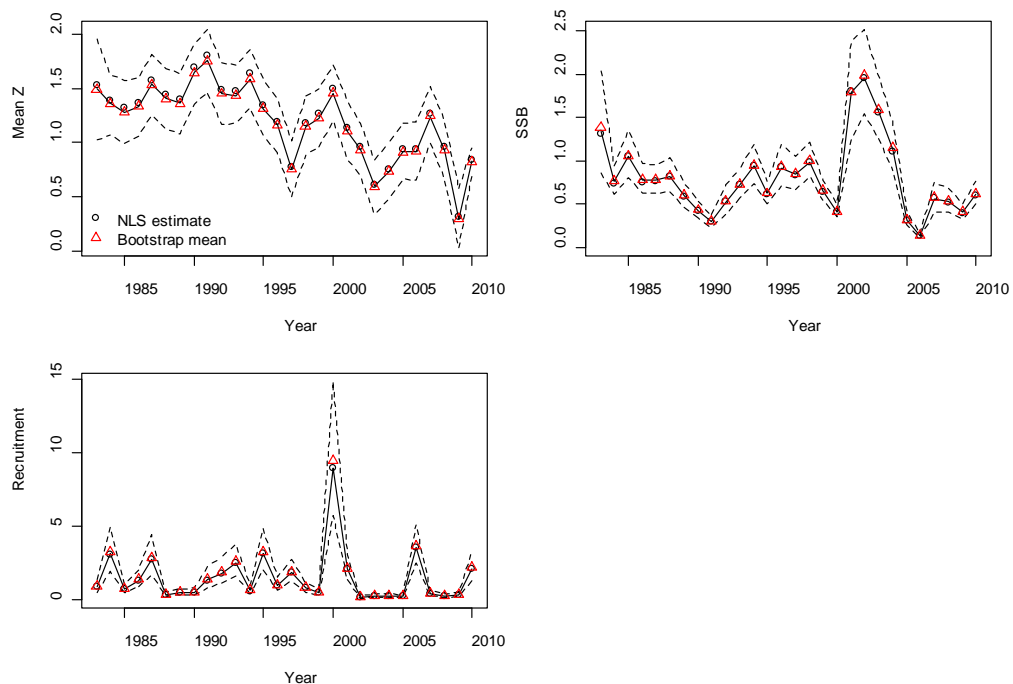


Figure 2.10.1. Sample stock output from SURBAR, showing model fits for North Sea haddock using the Scottish Q3 groundfish survey indices. Dotted lines give 90% confidence intervals.

Rapporteur's summary of presentation

This is an example of a model that is not general in the types of data that can be used, but is general in the number of stocks that have used it. This is a specific survey based model which can be biomass based but requires at least one age-structured survey to use.

Background: Started using this model because advice was required for TAC and effort based management. Additionally, TACs were becoming more restrictive due to fal-

ling stock size which was leading to more stringent advice. In early 2000s there was increasing concern with misreporting and black landings. Management is a bit stricter now and the data are more reliable. But there were certainly several years when they had serious concerns about the quality of catch data. This is one of the main reasons for being interested in survey based assessments.

Northern Shelf Working Group: At one of these meetings they started with a model originally made by Robin Cook. Then they spent quite a lot of time at the meeting working on the early versions of the SURBA model. After all of this, it seemed reasonable to start to base advice on surveys.

Other things the survey based assessment will help with:

- 1) When you have early stock estimates from the current year. For example, when your survey takes place before a working group, can use this model then use catch later to modify for the next year.
- 2) Useful for developing abundance estimates for smaller areas than management units. Difficult to disaggregate catch data into north and south (e.g. north and south whiting), but can do this with the survey.

Drawbacks: Mortality estimates can be very noisy. The surveys are only out for a small portion of the year, and they are general biomass sampling surveys which are not really aimed at a particular species. These might not be the most representative samples for a particular species. Additionally, the survey method does require assumptions about catchability, which you need to input as an external parameter. The model may not fit in to management paradigm if an absolute abundance estimate is required. However, if management is all about trend analysis then SURBA might be more appropriate.

Key assumption in this model is that mortality is separable and that cohorts decline exponentially. Need to estimate recruitment when cohort first enters population, and the two components of mortality as go through time. Parameters to estimate are R , F , S , and as you get more years, you get more parameters. Fixed values are year effect, age effect, and reference age. Important to consider how good these fixed estimates are.

SURBA Method: Must have a fixed value of q because it is very difficult to get estimation of this. All information starts at the beginning of year, then end up with SSQ for age structure indices and biomass indices. The smoothing penalty term (originally from R. Cook) is a way to penalize large interannual variations in the year effect on total mortality. Weight on this is determined by user and there is really no good way to estimate this value. So there are a number of ad-hoc choices made for this method.

Uncertainty Estimation: application of delta methods in SURBA 3.0, in SURBAR parameter resampling from multivariate normal based Hessian, much more straightforward.

Requirements: one age-structured survey, biomass, stock weights, maturity, natural mortality, series of user defined parameters.

Implementation: Fortran 90 code, has NAG Fortran 90 library. This is a driver for developing "minpack.lm" library to allow for distribution of SURBAR. Library is the same as what is used for CSA. Noel is also working on a number of distributions and flavours of this method.

Known problems: Provides relative abundance estimates only and requires assumptions for final year and oldest age. Model also assumes separability and mortality

uncertainty estimates can be very wide. Additionally, the R version is in the very early days of development.

Coby next presented an example of North Sea haddock

Windows Interface: Choose which index to go into analysis. Interface is limited to 30 surveys. Then set smoothing parameters (difficult to determine), reference age, etc. Retrospective analysis is also available.

Ease of Use: Relatively easy to use as model uses same VPA file formats already used for ICES assessments. May not be best method for what is available. Interface has options for plotting input data and model fits of SSB, recruitment, total biomass, total mortality (mean). Coby then showed plot examples of total mortality, residuals, retrospective analysis. Method is similar to CSA where comparing surveys with each other, but catch staying the same. Rarely get a pattern here, but will get a retrospective pattern in mortality. ScoGES provides uncertainty estimates.

One problem have had in past when using this model is when population signals are not that strong from one survey to another such as what happened with North Sea cod. Coby showed survey plot.

This model is good for situations when have good survey data and poor catch data. It does not perform well for flatfish stocks in North Sea. Another potential use of this method is in an exploratory sense just to see what the survey is doing.

2.10.3 Summary of WKADSAM discussion

Questions on where to put q's and if there was catch in this model.

Coby: No catch

Doug: Could you use this [catch] data as a constraint? May be getting realizations which imply a catch which is less than you know is the minimum taken. In this model you have an objective function where you could add some other terms such as "Given an M, the implied loss to fishing mortality must be at least a min value."

Coby: another option is to use the age composition data from catches. Level of catch is not used in assessment because the assumption is that tonnage is wrong. But the age comp is still representative of the population. This would also tie down the estimates.

Doug: also consider if you have a situation where selectivity pattern is different between what is and is not reported. Can have a situation where large vessels are fishing further offshore with observers then smaller vessels which do not have observers and are operating in different areas. This will influence the age comps.

Coby: this may not be a problem now because of stricter management.

Benoit/Coby: Exploration of harvest control rules are based on survey designs, so should have some methods in toolbox which do not rely on catch. This may be useful when the fishery starts to close.

Noel: so using this model when catch data becomes unreliable.

A discussion followed about fixing one of the age effects to 1 and why it mattered what age was chosen. The problem may be related to the amount of information available for estimation on the age chosen. The variance of the fixed age translates to the other ages so if estimation is poor than reflected through the rest of the ages. This is similar to standardizing cpue where it is best to pick factors that have large

amounts of information to standardize. Another issue may be the large number of surveys used.

Another discussion occurred on how to treat the q 's at each age for the survey. Have to fix one q and may be able to estimate the other q 's for other surveys from this. Can also have other information about relative catchability at age, and this is more useful than trying to estimate q inside the model.

NOTE: Little fuzzy on the discussion after this, but general notes are below.

Doug: how viable is the separability assumption? Saying M is varying the same way fishing mortality is varying at age. Specify q separately for every age group. Problem when have M on same magnitude of F . Normally think of s as a relative q .

Coby: assumption is that changes balance out between M and F , and that may be too large of an assumption. But q is not a mortality, it's a proportionality constant for a given survey. But the total mortality s is different. There is an option in program to put in M , but somewhat meaningless since have no information on this.

Doug: why can't you use this for TAC advice? Why can't use for status quo advice? This is perfectly reasonable as a control rule. Relative F would be just fine.

Coby: could use this when do have evidence of change, but what about something like North Sea cod when mortality estimates are not changing

Discussion continued on how to provide TAC advice without using catch data especially given the pressure to include fisheries data and the potential situation where the surveys have high variance and commercial data are improving. Situation is advancing and will try to use these data more in the future.

2.11 XSA

2.11.1 Description

No description of the model package was provided by the presenter.

Rapporteur's summary of presentation

Extended Survivors Analysis (XSA) is a very old program and is based on the old terminal numbers (survivors) of a cohort. Is iteratively tuned to cpue data and does not require any non-linear optimization algorithms to solve the system of equations.

Default options include time-series weighting, recruitment regression model, minimum number of data points, shrinkage for more robust regression (P shrinkage) and F shrinkage to aid with retrospective problems (add in the mean to improve robustness of prediction), the default value of F shrinkage is 0.5. Standard error thresholds and fleet weights, which can be tuned based on the residuals patterns in the data. Very similar to some of the features in ADAPT. Uses age-based indices, not aggregated indices, for tuning.

2.11.2 Summary of WKADSAM discussion

See Section 2.12.2 below.

2.12 B-ADAPT

2.12.1 Description

No description of the model package was provided by the presenter.

Rapporteur's summary of presentation

B-ADAPT is a modified version of ADAPT that is used to estimate unallocated mortality (miss-reporting of the catch), or a catch multiplier. It was designed specifically for the North Sea cod stock to deal with the known catch miss-reporting problem. In practice with standard VPAs there were problems with retrospective bias, and apparently the surveys were getting better at catching fish over time (i.e. q is apparently increasing). To solve this problem in the retrospectives, B-ADAPT was developed with the additional flexibility of including a catch multiplier. Simulated data were used to demonstrate that these multipliers could be estimable if a constraint on catch is used.

2.12.2 Summary of WKADSAM discussion

Following the presentations on the two models there were some general discussions on the following:

- The modification to ADAPT to include the catch multiplier addresses the previous retrospective patterns for the North Sea Cod.
- The modification is also similar to alternative explanations in the residual patterns such as time varying estimated via a constrained random walk.
- The method assumes constant q in the relative abundance index so there is some concern that the catch multipliers would vary erratically over time do noise (observation errors). In the case of the North Sea cod, industry openly admitted what catch they were taking, so this was used to constrain the misallocation of the catch.
- Anders State-space random effects model has the same catch multiplier and the results between the two methods are remarkably similar. The real concern with B-ADPAT suffers if there are year affects in the data.
- The catch multipliers appear to be fairly stable; as new data accumulate the estimates of the catch multipliers do not appear to vary too much. However, if there are real changes in q over time in the survey data, then this would cause changes in the catch multipliers and create a retrospective problem.
- Comments were made on the VPA backwards convergence properties and the differences that would be obtained in an SCAA. Another important point is that there must be a contrast or change over time in the catch reporting rate in order to properly estimate these catch multipliers (at least in the SCAA case).
- Overall, catch misreporting is a challenging subject for all assessments because all stock assessment models (VPA or SCAA types) are conditioned in one-way or another by the historical catch data (either conditioning the model on historical catch or fitting the models to historical catch). There are ways that SS can be tricked (e.g. a phantom IUU fleet) to estimate this unallocated catch. Multifan can predict missing years of catch or effort based current estimates of q for a given fleet.

2.13 General issues in benchmarks

2.13.1 Summary of presentation

According to the general Terms of Reference of ICES regarding benchmarks, those workshops aim at:

- evaluating the appropriateness of stock assessment data and methods;
- agreeing and documenting preferred stock assessment method for each stock;
- updating the relevant Stock Annexes to include what WS participants identify as current best practice assessment inputs and methods, providing sufficient detail to ensure that assessments can be readily replicated.

30 stocks have been reviewed during the period 2009–2010 through 6 thematical benchmarks: 2 on roundfish stocks (WKROUND 2009–2010), flatfish (WKFLAT 2009–2010), short-lived species (WKSHORT 2009), deepwater stocks (WKDEEP 2010). More benchmarks are already scheduled for the next two years.

Benchmark workshops are in their learning curve for the ICES community and still look for the best format to make this exercise the most efficient as possible. The distinction between assessment working groups and benchmark has not always been clear for participants during the last two years but it is expected that the benchmark process will be a natural element as more and more stocks will be benchmarked in the years to come.

What is a benchmark workshop? For consistency between years, assessment working groups within ICES are expected to follow a protocol described in a document called 'stock annex'. Stock annex is often described as the recipe of a stock assessment with information on the adopted model to use, the value of each parameters. Assessment working groups use those information with updated datasets relevant to the fisheries and surveys but are not supposed to switch to other models or parameters except for exploratory purposes. A benchmark workshop will review the available informations and methods used within the assessment working group. The final piece of work during those meetings is the writing or update of the stock annexes.

Review of the benchmark reports

Each stock in the 6 available reports has been reviewed with methodology in mind. The idea behind that was to identify stock by stock what has been changed and its implication for assessment models. The following items have been compared before and after the benchmark:

- Used, proposed and adopted models
- Issues with commercial catch and biological data
- Issues with commercial fleet and survey indices
- Issues with retrospective patterns
- Changes in recruitment, forecast and biological reference points (including MSY estimates).
- Recommendations from the benchmark participants regarding methods.

The intent of this review was not to go into details for each stock as benchmarks had already identified and addressed (or not) their issues but to identify common issues among the various stocks. A summary table has been built for the whole set of stocks. If some issues have been identified, the cell is coloured. When discussions on those issues were inconclusive or no solution was proposed, the cell is in yellow. This effectively led to no change on that part in the final stock annex. If the changes have resulted into improvement for the assessment, the cell is green. In the case of an unsolved critical issue, the cell was red. In that case, the previous assessment methodology was rejected and no new protocol is suitable to properly evaluate the stock.

It is important to note that this work only represents a portion of the ICES stocks and the benchmark process in general is still in its learning curve within ICES.

Change in stock assessment models

Among the 30 stocks, prior to the benchmark, about two thirds used to be assessed using VPA based approaches including XSA and B-Adapt. Few new models have been introduced during the workshops but most stocks have had the opportunity to test alternate models. The level of testing has been very variable depending on time available, effort and interest to learn how to set and run those models.

Table 12.13.1. Change in stock assessment models before and after benchmarks. Cells coloured in yellow mean that some proposals for new models/methods have been made but were not accepted. Cells in green means the change had a positive impact on some assessment issues for a given stock. Cells in red means the assessment has been rejected because of a problem (generally data-related).

Stock	Assessment models		
	Before	Proposed during benchmark	Adopted
Whiting VIlId-IV	XSA	-	XSA
North Sea Cod	B-Adapt	SAM	B-Adapt
Celtic Sea Cod	XSA	SAM, B-Adapt	no assessment
Cod in IIIa East (Kattegat)	B-Adapt	SAM	SAM
Western Baltic Cod	XSA	SAM	SAM
Eastern Baltic Cod	XSA	SAM	XSA
Northeast arctic saithe	XSA	-	XSA
Saithe in Icelandic waters	ADCAM	C-at-age ADMB	C-at-age ADMB
Faroe saithe	XSA	Adapt	XSA
Northern Hake	XSA	SS3	SS3
Southern Hake	VPA based model	Gadget	Gadget
Sole VIIe	XSA	XSA, AM	no assessment
Sole VIId	XSA	-	XSA
North Sea Plaice	XSA	stat. c@age	XSA (CSA for exploration)
Plaice VIId	XSA	-	XSA
Plaice VIIe	XSA	SS3	XSA
Sole in IIIa	XSA	SAM	SAM
North Sea Sole	XSA	SAM, stat catch@age	XSA
Roundnose Grenadier Vb,	VPA based model	c-at-age, surplus, LPUE	c-at-age, surplus, LPUE
Greater forkbeard	VPA based model	Stock depletion model	Stock depletion model
Tusk in Va and XIV	trends in landings/survey	Gadget	Gadget
Portuguese dogfish	none	Bayesian demographic model	Bayesian demographic model
Leafscale Gulper shark	none	use of indicators	indicators, surplus
Red sea bream in X	ad-hoc VPA, XSA	analysis of trends of abundance	analysis of trends of abundance
Greater silver smelt in Va	none	Coleraine	trends on survey
Greater silver smelt in other	XSA	XSA	trends on survey
Barents Sea Capelin	CapTool Bifrost	-	CapTool Bifrost
Bay of Biscay anchovy	BBM	Brem	BBM
Icelandic Capelin	backcalculation of cohorts	-	no assessment
Sprat in Subarea IV	CSA (abandoned in 2009)	length based model	no assessment

VPA-based models are used in one-third of the 30 stocks benchmarked by the end of 2010, and more diversity in terms of models has been introduced through the benchmarks. Stock-specific models are rarely used. Some stocks no longer have any proper assessment. This is generally a data-related problem that no model during the benchmark has been able to solve.

Table 12.13.2. Evolution of the repartition of the type of models before and after the benchmarks.

Models	Prior 2009-2010 benchmarks	Situation late 2010
XSA	16 stocks	9
B-Adapt	2	1
VPA	4	0
SAM	0	3
Gadget	0	2
SS3	0	1
Bayesian prod model	1	4
Indicators / Trends	2	5
Stock specific model	1	2
Other	0	1
no assessment	3	4

Model selection is mainly done based upon the types of available datasets but also on additional consideration such as the ease of the transition to another model.

Training, reformatting datasets and parameters and testing are time consuming processes. The help of the creators of those models prior and/or at the benchmark speeds up the transition to other models. Models need to be sufficiently documented. Participants also requested workshops to learn how to use and how to consider new models prior to the benchmark. Some comments have been made about having a 2- or 3- stages benchmark with a first meeting to discuss problems and to consider some solutions to test and a final meeting reviewing the results and adopting a new methodology.

The transition to new models has itself room for development in model evaluation methodology. Some benchmark participants pointed the lack of protocol within ICES for the evaluation and acceptance of new stock assessment models (e.g. "what to look for? what is acceptable?"). Other members have requested some protocols and tools in order to be able to compare substantially different models (e.g. age-based models vs. surplus production model).

Finally, other members believe indicators can also be informative of the status of a stock especially in the case of data-poor stocks where an analytic assessment if doable may be only used as indicative of trends. Many deepwater stocks reviewed during WKDEEP in 2010 are using indicators such as abundance indices derived from log-books and surveys to provide assessment indicative of trends.

Issues with commercial catch and biological data

As model selection is driven by the availability of suitable datasets, the second step of this review was to consider problems with the data.

Landings and discards are expected to be the most easily available datasets (especially landings) but several stocks experience mis- or underreporting. Misreporting is a serious issue and has led to reject the assessment methodology for Celtic sea cod in 2009. Other assessments have been considered as only indicative of trends because of the lack of documented discards or gaps in time-series. There are demands for methods to properly reconstruct gapped time-series.

Age and growth are a problem for some stocks when ageing has been found to be inaccurate. Before the first benchmarks, most stocks used to rely on age-based models. The problems highlighted with new data, otolith exchange has led some stock like northern and southern hake to abandon the traditional age-structured approach for length structured models. This however still relies on a proper assumption.

tion/information on growth at some point and in some case, it is still insufficient. Those stocks then rely on other approach such as surplus production models for example which only depend on landings and abundance indices/cpues.

Maturity does not appear to be a problem for stock assessment although sensitivity analyses have shown on some stocks that models outputs were substantially sensitive to the values of the ogives. Some revisions have been done with mixed effect depending on the stock.

Estimating *Natural Mortality* remains a challenge in stock assessment. For most benchmarked stocks, no attempt has been made to recalculate this parameter as it is generally considered that fishing mortality is much higher than natural mortality. For North Sea cod, an annually variable mortality parameter has been introduced. For short-lived species, this parameter may be equal to or higher than fishing pressure and is often dependant on external factors such as climate, trophic conditions and predator/prey interactions. WKSHORT tried to review those aspects but so far no model using time-series of environmental factor is operational or able to explain the variation of mortality.

Table 12.13.3. Issues with commercial catch and biological data.

Stock	Adopted model	Catch and Biological Data issues				
		Landings	Discards	Age/Growth	Mortality	Maturity
Whiting VId-IV	XSA	no	no	no	no	no
North Sea Cod	B-Adapt	no	no	no	annual var ?	annual var ?
Celtic Sea Cod	none	underreporting	highgrading	no	no	new ogive
Cod in Illa East (Kattegat)	SAM	uncertain	not included	no	no	no
Western Baltic Cod	SAM	no	no	w-at-age	no	new ogive
Eastern Baltic Cod	XSA	underreporting	poor estimates	age reading, w-at-age	no	new ogive
Northeast arctic saithe	XSA	no	no	no	no	no
Saithe in Icelandic waters	C-at-age ADMB	no	no	no	no	no
Faroe saithe	XSA	no	no	updated catch@age	no	no
Northern Hake	SS3	no	no	overestimation of age	higher ?	no
Southern Hake	Gadget	no	no	overestimation of age	higher ?	new ogive
Sole Vile	none	misreporting	no	no	no	no
Sole Vild	XSA	no	no	no	no	no
North Sea Plaice	XSA (CSA for exploration)	no	uncertainties	no	no	no
Plaice Vild	XSA	no	gaps	no	no	no
Plaice Vile	XSA	no	exp. Low discards	no	no	no
Sole in Illa	SAM	no	no	no	no	no
North Sea Sole	XSA	no	not included	no	no	no
Roundnose Grenadier	c-at-age, surplus, LPUE	misreporting	substantial	age reading (got rid of)	no	no
Greater forkbeard	Stock depletion model	partial	partial	no	no	no
Tusk in Va and XIV	Gadget	no	no	faster growth obsd	change in M	no
Portuguese dogfish	Bayesian demographic model	combined species	no	no	no	no
Leafscale Gulper shark	indicators, surplus	combined species	no	no	no	no
Red sea bream in X	analysis of trends of abundance	no	no	no	no	no
Greater silver smelt in Va	trends on survey	no	no	shift in c@age	no	no
Greater silver smelt other	trends on survey	no	no	no	no	no
Barents Sea Capelin	CapTool Bifrost	no	no	no	no	no
Bay of Biscay anchovy	BBM	no	no	no	link with enviro	no
Icelandic Capelin	none	no	no	no	too low M	no
Sprat in Subarea IV	none	no	no	no	no	no

Issues with commercial fleet and surveys indices

For most stocks, data from tuning fleets and survey indices are part of the datasets used for the stock assessment. For some other stock, the lack of proper data led the stock to be assessed only through trends and as a consequence, those indices are the only output of the stock assessment.

During the benchmarks, substantial work has been accomplished to improve the existing datasets. This involved manipulating or splitting surveys, recalculation based on new data. Several benchmarks requested development in methods of standardization of commercial cpues in order to take account factors such as seasonal patterns, technological changes. There were also requests for methods to integrate multiple tuning indices especially from different subareas within a single-stock unit.

Further studies are also requested on the impact of missing/combined indices over an assessment.

Table 12.13.4. Issues with indices from commercial fleets and surveys.

Stock	Adopted model	Indices issues	
		Commercial fleet	Survey
Whiting VId-IV	XSA	Opposite trends	Opposite trends
North Sea Cod	B-Adapt	no	extended indices
Celtic Sea Cod	none	no	combined indices
Cod in IIIa East (Kattegat)	SAM	no	no
Western Baltic Cod	SAM	new tuning fleet	larger spatial overlap needed
Eastern Baltic Cod	XSA	inconsistencies in age comp.	
Northeast arctic saithe	XSA	no	no
Saithe in Icelandic waters	C-at-age ADMB	CPUE data not standardized and included	no
Faroe saithe	XSA	revision of indices	high CV (not included for assessment)
Northern Hake	SS3	no	no
Southern Hake	Gadget	no	no
Sole VIIe	none	no	limited survey coverage
Sole VId	XSA	HP correction for Belgian trawl tuning flt	combined survey indices
North Sea Plaice	XSA (CSA for explor)	standardized lpue series	combined survey indices
Plaice VId	XSA	HP correction for Belgian trawl tuning flt	combined survey indices
Plaice VIIe	XSA	no	no
Sole in IIIa	SAM	review of tuning fleets	directed survey
North Sea Sole	XSA	technical change / correction in CPUE	updated database
Roundnose Grenadier Vb, VI, VII	c-at-age, surplus, L	no	no
Greater forkbeard	Stock depletion m	no	no
Tusk in Va and XIV	Gadget	no	no
Portuguese dogfish	Bayesian demogra	combined data, LPUE survey	lack of regular survey
Leafscale Gulper shark	indicators, surplus	combined data, LPUE survey	lack of regular survey
Red sea bream in X	analysis of trends	no	no
Greater silver smelt in Va	trends on survey	new glm model (no proper fit)	new method for BT survey
Greater silver smelt in other areas	trends on survey	no	no
Barents Sea Capelin	CapTool Bifrost	no	no
Bay of Biscay anchovy	BBM	no	DEPM parameter, coverage in coastal area
Icelandic Capelin	none	no	no
Sprat in Subarea IV	none	no	large infrequent hauls => bias

Retrospective patterns

For 10 benchmarked stocks, there have been efforts to reduce retrospective patterns. Those efforts included revision of plus group composition, revision and splitting of tuning series, reconstruction of catch-at-age, discards and taking account of migration. No solution was found for some stock and it eventually led for Sole in VIIe to the rejection of the assessment methodology.

The evaluation of retrospective patterns lacks of a proper metrics therefore the development of diagnostic tools for the measurement of those patterns has been recommended as this procedure should be routinely applied to any stock assessment.

Table 12.13.5. Issues with retrospective patterns and proposed solutions to reduce them.

Stock	Adopted model	Retrospective patterns	
		present ?	revision
Whiting VIIId-IV	XSA	not invest.	
North Sea Cod	B-Adapt	not invest.	
Celtic Sea Cod	none	not invest.	
Cod in IIIa East (Kattegat)	SAM	not invest.	
Western Baltic Cod	SAM	substantial	unsolved
Eastern Baltic Cod	XSA	not invest.	
Northeast arctic saithe	XSA	reduced	plusgroup, splitting tuning series, less shrinkage, no tapering
Saithe in Icelandic waters	C-at-age ADMB	reduced	use of statistical c-at-age model
Faroe saithe	XSA	reduced	catch-at-age, commercial tuning series
Northern Hake	SS3	not invest.	
Southern Hake	Gadget	not invest.	
Sole VIIe	none	strong	unsolved
Sole VIIId	XSA	not invest.	
North Sea Plaice	XSA (CSA for exploration)	not invest.	
Plaice VIIId	XSA	reduced	discards/migration
Plaice VIIe	XSA	reduced	migration
Sole in IIIa	SAM	substantial	unsolved
North Sea Sole	XSA	reduced	commercial tuning series
Roundnose Grenadier Vb,	C-at-age, surplus, LPUE	not invest.	
Greater forkbeard	Stock depletion model	not invest.	
Tusk in Va and XIV	Gadget	not invest.	
Portuguese dogfish	Bayesian demographic mo	not invest.	
Leafscale Gulper shark	indicators, surplus	not invest.	
Red sea bream in X	analysis of trends of abund	not invest.	
Greater silver smelt in Va	trends on survey	not invest.	
Greater silver smelt in other	trends on survey	not invest.	
Barents Sea Capelin	CapTool Bifrost	not invest.	
Bay of Biscay anchovy	BBM	moderate	unsolved
Icelandic Capelin	none	not invest.	
Sprat in Subarea IV	none	not invest.	

Recruitments and forecasts

Few attempts have been made to provide substantial change on methods regarding recruitments and forecasts.

Substantial research are currently done on linking recruitment with environmental drivers for some stocks in particular short-lived species which is a major challenge for forecast and management for those stocks. The state-of-the-art regarding the interactions between those stocks, environmental and trophic conditions has been presented during the WKSHORT benchmark (2009) but no model is currently able to provide any forecast based on those factors.

Table 12.13.6. Changes in methods for the evaluation of recruitment and forecast.

Stock	Adopted model	Changes in recruit. estim.	Changes in forecast
Whiting VIId-IV	XSA	not invest.	not invest.
North Sea Cod	B-Adapt	not invest.	not invest.
Celtic Sea Cod	none	none	none
Cod in IIIa East (Kattegat)	SAM	not invest.	not invest.
Western Baltic Cod	SAM	not invest.	not invest.
Eastern Baltic Cod	XSA	not invest.	not invest.
Northeast arctic saithe	XSA	revised	revised
Saithe in Icelandic waters	C-at-age ADMB	revised	revised
Faroe saithe	XSA	revised	revised
Northern Hake	SS3	revised	revised
Southern Hake	Gadget	revised	revised
Sole VIIe	none	none	none
Sole VIId	XSA	variable recr	not invest.
North Sea Plaice	XSA (CSA for exploration)	not invest.	not invest.
Plaice VIId	XSA	not invest.	not invest.
Plaice VIIe	XSA	revised	unchanged
Sole in IIIa	SAM	not invest.	not invest.
North Sea Sole	XSA	not invest.	not invest.
Roundnose Grenadier Vb, Vc	C-at-age, surplus, LPUE	not invest.	not invest.
Greater forkbeard	Stock depletion model	not invest.	not invest.
Tusk in Va and XIV	Gadget	in devlt	in devlt
Portuguese dogfish	Bayesian demographic model	not invest.	not invest.
Leafscale Gulper shark	indicators, surplus	not invest.	not invest.
Red sea bream in X	analysis of trends of abundance	not invest.	not invest.
Greater silver smelt in Va	trends on survey	not invest.	not invest.
Greater silver smelt in other	trends on survey	not invest.	not invest.
Barents Sea Capelin	CapTool Bifrost	unchanged	unchanged
Bay of Biscay anchovy	BBM	link with envir	link with envir
Icelandic Capelin	none	not invest.	none
Sprat in Subarea IV	none	none	none

Uncertainties and Biological Reference Points

ICES is currently implementing a MSY approach for all stocks since the beginning of 2010. Therefore all previous benchmarks might not have considered this new approach but rather the existing biological reference points. Upcoming benchmarks are likely to have more focus on MSY indicators. MSY has not been strongly investigated through the benchmarks so far. Some assessment working groups have done this exercise in parallel to benchmarks following procedures written during WKFRAME (2010). Few revisions have been made during the benchmark workshops on biological reference points.

Twelve benchmarked stocks use models providing estimates of uncertainties. Some benchmark participants have recommended some work to be done on how those uncertainties could be integrated into advices. Bay of Biscay anchovy already provides annual advices with catch options including probabilities to have SSB below Blim which basically introduces the notion of risk associated with a level of catch.

There have been some attempts to provide estimates of reference points following the MSY approach (WKROUND, 2010). At least 3 stocks are currently assessed with models which are structurally able to provide MSY reference indicators on an annual basis with probabilities for biomass or fishing mortalities to be below those indica-

tors. This is different from the ICES approach which suggests a long-term (i.e. constant) target F_{MSY} .

Table 12.13.7. Availability of estimates of uncertainties in model outputs and changes in reference points.

Stock	Adopted model	estimates of uncertainties	Revised Reference points			
			Blim, Bpa	Flim, Fpa	MSY approach	Fmax,0.1,msy
Whiting Vllid-IV	XSA	no	not invest.	not invest.	not invest.	not invest.
North Sea Cod	B-Adapt	yes	unchanged	unchanged	not invest.	revised
Celtic Sea Cod	none	no	not invest.	not invest.	not invest.	not invest.
Cod in IIIa East (Kattegat)	SAM	yes	revised	removed	not invest.	revised
Western Baltic Cod	SAM	yes	unchanged	unchanged	not invest.	unchanged
Eastern Baltic Cod	XSA	no	removed	removed	not invest.	not invest.
Northeast arctic saithe	XSA	no	unchanged	unchanged	MSY estim.	revised
Saithe in Icelandic waters	C-at-age ADMB	no	not invest.	not invest.	MSY estim.	revised
Faroe saithe	XSA	no	revised	revised	MSY estim.	revised
Northern Hake	SS3	yes	revised	revised	MSY estim.	revised
Southern Hake	Gadget	no	unchanged	unchanged	MSY estim.	revised
Sole Vlle	none	no	removed	removed	not invest.	not invest.
Sole Vlld	XSA	no	unchanged	unchanged	not invest.	not invest.
North Sea Plaice	XSA (CSA for exploration)	no	not invest.	not invest.	not invest.	not invest.
Plaice Vlld	XSA	no	not invest.	not invest.	not invest.	not invest.
Plaice Vlle	XSA	no	not invest.	not invest.	unconclusive	revised
Sole in IIIa	SAM	yes	not invest.	not invest.	MSY estim.	revised
North Sea Sole	XSA	no	not invest.	not invest.	not invest.	revised
Roundnose Grenadier Vb,	C-at-age, surplus, LPUE	yes	not invest.	not invest.	MSY output	not invest.
Greater forkbeard	Stock depletion model	yes	not invest.	not invest.	not invest.	not invest.
Tusk in Va and XIV	Gadget	no	not invest.	not invest.	not invest.	not invest.
Portuguese dogfish	Bayesian demographic mo	yes	not invest.	not invest.	MSY output	not invest.
Leafscale Gulper shark	indicators, surplus	no	not invest.	not invest.	not invest.	not invest.
Red sea bream in X	analysis of trends of abund	no	not invest.	not invest.	not invest.	not invest.
Greater silver smelt in Va	trends on survey	no	not invest.	not invest.	not invest.	not invest.
Greater silver smelt in othe	trends on survey	no	not invest.	not invest.	not invest.	not invest.
Barents Sea Capelin	CapTool Bifrost	yes	not invest.	not invest.	MSY output	not invest.
Bay of Biscay anchovy	BBM	yes	not invest.	not invest.	not invest.	not invest.
Icelandic Capelin	none	no	not set	not invest.	not invest.	not invest.
Sprat in Subarea IV	none	yes	not invest.	not invest.	not invest.	not invest.

Summary of recommended work

The following is a quick summary of the main recommendations made by the different benchmark workshops:

- 1) Use of stock assessment models
 - Development of protocols to evaluate and select the most appropriate model considering the availability/quality of datasets
 - Development for some stocks of models with linkage with environmental factors and other species.
 - Assessment scientists require some framework for training (workshops, tutorials, documentation)
- 2) Data and parameters
 - Development of methods to fill gaps in time-series
 - Methods to manipulate, standardize cpues and survey indices
- 3) Assessment outputs
 - Development of methods/metrics to evaluate retrospective patterns
 - How to integrate uncertainties into advice?

Rapporteur's summary of presentation

Analysis of the benchmarking process to investigate what issues were identified that suggests a change in methodology might be required, what alternatives were investigated, and what outcome resulted. The benchmarking resulted in an increased diversity of models being used, and several stocks being given no assessment.

A variety of issues arose during the evaluating and transition phase:

- How to compare different kinds of models
- How to manage the transition between one model and another
- How to deal with data problems including survey and catch history
- How to deal with model problems, specifically retrospective trends.
- How to incorporate uncertainties in MSY estimates.

2.13.2 Summary of WKADSAM discussion

- Length data could have enough model structure to be useful, which means length-based models could be usefully used where age-based methods struggle because of poor age information.
- Simply making this report available to working groups may be helpful to give working groups an idea of what possibilities are available.
- How much pressure is there from MSC to adopt an MSY approach? ICES is still working out where it stands on this and starting to work towards achieving Fmsy by 2015. ICES may have headed in this direction without understanding fully what it implies. ICES has to head in this direction because of commercial pressure to achieve MSC certification. Should there be informal links with MSC to ensure consistency of approaches?
- Why have Northern and Southern Hake decided to take different approaches? Problems believing aging suggested a length-based approach should be used. Perception (incorrectly) that Multifan-SL couldn't use abundance indices (there is way of incorporating it by defining it as another fishing fleet). Gadget perceived to be too complicated (built as an ecosystem model, but can also run in single species mode). Stock synthesis is already well tested to an extent that couldn't be achieved with a custom model in a limited time. Several models may have been appropriate, so in the end it may come down to access to relevant expertise.
- Are models being rejected because they are too stiff? If errors in catch aren't included F seems to be too variable. Post hoc smoothing of F can be applied.
- Is the choice of what to do based on personality within the working group or objective criteria?
- Users need to look at diagnostics and residuals and think about whether the data or model assumptions are wrong.
- Should ICES be more involved in how a suitable model is selected? Should there be a list of formal requirements. There is no sheet with requirements about how the model should be selected. The latest guidelines on benchmarking include some guidance, but not very clear.
- What about where there's a vague sense of unease about the model being used? Consistency should be valued above small improvements in model accuracy.

- There are a variety of SCAA approaches under development. How do we work towards having a more consolidated approach? Is it the time to do this?
- Why didn't some stocks switch from XSA to SAM despite considering the latter? Personal experience – Anders Nielsen was the only person in the group who had experience with the model, so the default behaviour is often to stick with a model that one has most experience with. We need people to get experience with different models.
- How much were the considered models actually investigated, or was it just a brief experiment? It's hard to look at the table presented and see what actually happens.
- It takes time to transition between models because of gaining expertise, time pressure means that the scientists are often cautious about embarking on a new method if they don't know that they have enough time to explore it appropriately. The process of considering available options (model packages, custom approaches), then eventually moving to Stock Synthesis for Northern hake took about 6 months, including time investigating the data requirements and becoming familiar with the model. Given this, there is a need to think about amount of preparatory time before the benchmark workshop, possibly a preparatory workshop, followed some time later by the benchmark workshop. It's a difficult decision, to adopt a new model, because there's no back up within the assessment working group if things go wrong. It also depends on relevant experts being available.
- How did learning curves for Gadget and Stock Synthesis compare? Similar
- When transferring between models, even where data are transferable between models, control structures are idiosyncratic and vary between models. This is partly mitigated by good documentation. It is important to have an expert because it's possible to misunderstand well-documented software and understand the interactions between the model parameters.
- The benchmark process is something in its infancy – there was only one or two months notice for the first benchmarks. The idea is to create consistency in the other assessments – but problems can arise from the data. We should be talking about how the process *should* look rather than how it currently looks, including looking at data before putting it into a model. There's an interesting question to be had about "crank turning" in the intervening years and how this relates to the MSE approach.
- What happens with problems in defining stock identity? This affects more management than assessment. Managers don't want to move management boundaries.
- Genetics has been rather a flop at identifying stocks, but could be useful for identifying close kin, which is useful for identifying whether stocks are discrete. Also it can be used for mark-recapture studies, generally suited to stocks with a small number of individuals.

2.14 Generic model features

2.14.1 Summary of presentation

The presentation offered discussion points on five topical issues in stock assessment.

2.14.1.1 Plus-group paucity

Catch-at-age data frequently reflect relatively few fish in the oldest age groups, which are often treated as a plus-group in assessments. VPA assessments making the standard assumptions for the value of natural mortality and of flat selectivity at the oldest ages interpret this as high cumulative fishing mortality along cohorts. However statistical catch-at-age (SCAA) approaches, which offer more by way of model fit diagnostics, can run into difficulties when making similar assumptions, as their results show systematic patterns which indicate that older fish should be caught than are observed. Examples of this include Gulf of Maine cod, New Zealand hoki, South African hake and Southern bluefin tuna, for which standard statistical model selection criteria reject assumptions of flat selectivity at large age coupled to conventional assumptions for natural mortality such as $M = 0.2$ independent of age.

Alternative assumptions which can remove this feature of the model fits include dome shaped selectivity, and higher values of M , particularly increasing M at larger ages. These in turn carry different implications for the values of standard reference points and appropriate scientific recommendations for management. However they can also give rise to reservations about the associated implications of substantial cryptic biomass, or an absence of plausible biological mechanisms to account for larger natural mortality at older ages (what could be eating these bigger older fish?).

Is frequent use of VPA in ICES (in particular) camouflaging this problem, which merits more attention?

2.14.1.2 Separate vs. Combined estimation

Typically the objective functions (usually the negative logs of penalised likelihoods) for SCAA assessments will include time-series of abundance indices, residuals about assumed stock–recruitment relationships, and proportion-at-age information for both commercial catches and surveys. More “statistically purist” approaches prefer to use data in the form they were originally collected, so that proportions-at-age are replaced by proportions-at-length together with a known growth curve, and may even extend to where the last is replaced by age–length key data with the growth curve parameters fitted in conjunction with the population dynamics model.

A particular problem with these approaches is how to assign appropriate relative weights (inverse variances) to these different data-types in objective functions, particularly when there is not the independence within each dataset which likelihood-based functions usually assume. Furthermore, while fitting directly to age–length key information can remove some biases in estimating the values of growth curve parameters, it may not prove very robust to the presence of biases in some of the ageing data.

It is desirable to attempt to model correlation explicitly to obtain more defensible likelihood functions, and hence obtain more realistic estimates of assessment precision. An example was given of Greenland halibut where incorporating AR1 processes in models for both age and year effects removes much of the systematic pattern otherwise evident in residuals for the fits to proportions-at-age data. Though introducing these features makes little difference to point estimates for past abundance trends in this case, they become important to include when moving on to approaches such as Management Strategy Evaluation (MSE) which require possible future data to be generated which display realistic levels of observation error.

2.14.1.3 Modelling catchability when selectivity-at-age varies over time

Since catchability q and selectivity at age S_a are aliased in relating catch rate to abundance, a frequent convention is to fix the maximum S_a value to be 1 so that q is uniquely defined. However problems arise in SCAA assessments which admit variations over time in S_a . If a dome-shaped S_a broadens, for example, it does not seem reasonable to assume that even if the underlying abundance is unchanged that the overall catch rate will nevertheless increase. Rather it seems more plausible to assume that catchability q will decrease in these circumstances.

This has important implications for drawing conclusions about trends in resource abundance from data giving evidence of time-varying selectivity, but it is unclear what the best approach is for adjusting q in concert with changes in S_a . It was suggested that choices among alternative possible approaches be advised by simulation testing, where the underlying mechanism generating time-dependence in S_a is a non-homogeneous distribution of the resource by age, with either this distribution, or the spatial pattern of fishing, varying over time.

2.14.1.4 Model selection

Assessment models including random effects, for example random walks in selectivity at age, are often (for reasons of time) being taken no further in practical application than the maximum penalised likelihood estimation stage. This renders use of AIC for model selection purposes problematic, as the random effects themselves are constrained rather than freely estimable parameters. One solution to this is Bayesian estimation combined with DIC, but this can be computationally infeasible in settings such as assessments needing to be conducted during the course of a (typically about one week long) international scientific working group meeting.

Another frequent problem in assessments is data conflicts, evidenced by the assignment of different relative weights to different data sources resulting in very different results. In this situation, the answer is not to seek the best relative weighting scheme within a single assessment approach, but rather to conduct separate assessments each using data that are not in conflict among themselves or with the model, then to consider the alternative results within some risk assessment framework (e.g. such as MSE).

One problem however is that data conflicts are not necessarily readily identifiable. An example was given of a standard VPA for Gulf of Maine cod with flat selectivity assumed for older ages to fix the fishing mortality on the plus-group. If some random variability was admitted in this relationship, once the associated variance was increased beyond a certain size, the solution jumped to an alternative trajectory reflecting higher biomasses. This is a situation possibly indicative of a multimodal objective function whose global minimum is sensitive to the relative weights accorded to different sources of information.

2.14.1.5 SCAA vs. VPA

SCAA would appear the better approach in principle. It is not constrained by possibly misleading backwards convergence, and offers better and statistically based approaches to examine mis-fitting in circumstances of, for example, retrospective patterns.

However there are competing practical considerations. If use of VPA may be somewhat akin to driving a car (relatively easy to learn and forgiving of mistakes), application of SCAA may be more like flying a helicopter (which is quite the opposite).

VPA's robustness properties may outweigh the potential advantages of SCAA unless appropriate expertise in use of the latter is available.

2.14.2 Summary of WKADSAM discussion

The presentation focused on key recurring problems seen in stock assessment and some possible solutions to these issues:

- 1) Plus Group Paucity. This refers to the relative lack of older fish that is actually seen than is expected from model predictions. In most stock assessments where flat topped selectivity is assumed, it results in higher abundance within the oldest ages or plus group than is actually seen in survey and catch data. Three possible solutions were presented including: 1. Using domed selectivity; 2. Increasing natural mortality; 3. An increasing natural mortality schedule with age. Each option results in significantly higher fit to Gulf of Maine cod data than the base case, which assumed flat selectivity and a constant natural mortality of $M=2$. However, none of the options were significantly better than the others and there was a relative lack of data to strongly support any option. In addition, domed selectivity results in large "cryptic" biomass not accessible to the fishery or surveys and higher M levels (≥ 4) might be pushing limits of believable M values for cod.

Discussion Points:

- What is the hypothesis behind the increasing M schedule option?
 - It is possible that old age is simply resulting in older fishing dying off and thus causing a higher M at these older ages. It is the best of the 3 options and avoids cryptic biomass.
- Is it possible that differences in Female/Male mortality rates might be causing this shift to increasing M -at-age? For US west coast rockfish it has been observed that females appear to have higher M at older ages.
 - This is a possible explanation that is still being looked into, however it is difficult to parse out this information from common fisheries data (i.e. there is a limit to how much information the data can give us).
 - Sex ratio shifts have been seen with Hoki where M increases during spawning season as mammals feed on spawning aggregations. Older fish appear to remain on spawning grounds for longer periods and hence face this higher mortality for an extended time period, which is a possible hypothesis supporting an increasing M schedule.
 - Southern Gulf of St Lawrence cod appear to show a similar trend of higher M 's due to seal predation. However, tagging studies with northern Gulf of St Lawrence cod show very low M (~ 1), but it appears that it is highly variable with high M 's seen 5–6 years ago.
- Is it possible that these suggested solutions are aliasing a problem elsewhere in the model since each solution has limited biological support?
 - It is possible and likely that the real problem might be difficulty in accurate ageing since fish become harder to age as they grow older. In reality, a mix of all 3 options might be a more palpable solution to the problem.
- It appears that changes in residual patterns are seen in the earlier years of the model for the older ages, but not in the later years. Why is this?

- It is likely an issue with the reliability of the data in the earlier years resulting in the biggest changes coming from this time period.
 - Is it possible that these changes in residual patterns are due to the assumption of constant selectivity?
 - It is doubtful due to the fact that these surveys (US Northeast Fisheries Science Center) are considered some of the most consistent over time.
- 2) Separate vs. Combined Estimation. The problem of how to weight multiple objective function components is becoming increasingly important, especially due to the fact that the number of terms within each component affects its influence resulting in non-independence. A solution to this problem is to either downweight components by multiplying by $1/(\text{\#of age classes})$ and/or to fit to age-aggregated indices and proportions-at-age instead of numbers-at-age.
 - 3) Catchability with Time-Varying Selectivity. Current SCAA allow year-varying selectivity, however cpue series rely on assumption of a constant catchability. If selectivity at a given age broadens, but q is constant it results in a higher catch for the same effort. This requires that selectivity-at-age must be renormalized or a constant average value of ages is maintained. Selectivity-at-age is likely constant over time, but the spatial distribution of the fishery changes resulting in higher catches in a given year of a given age because of the spatial age distribution of the species.

Discussion Points:

- Why wouldn't we believe that selectivity is changing over time?
 - Fishermen move to where the fish are in highest abundance and so if large age classes enter the fishery, then fishers will adjust to fish on these aggregations. The result is changes in spatial distributions of the catches and higher relative catches of a given age in a given year due to where the fishery is occurring, not in actually increased selectivity of the gear.
- 4) Model Selection. Often different datasets contain opposing signals regarding population trends, which result in model outputs being highly conditional on the relative weights given to the various input datasets used in the objective function. It is important to not "average" over these data conflicts, but instead to provide multiple model results representing different data signals (e.g. data weightings). Data conflicts often result in multimodal likelihood surfaces that can cause bifurcations in model results depending on what weighting is used.
 - 5) SCAA vs. VPA. SCAA is much more flexible than VPA, but VPA is generally easier to use and more forgiving to user mistakes. The key is to match the model to the situation. When expertise and available time are high, then it is appropriate to use SCAA and make custom built models. However, when available time and expertise is low, then VPA and generalized models might be more appropriate. For ICES assessments, where numerous stocks must be assessed in a limited time by a number of individuals with varying expertise, it is better to use a relatively simple VPA model that is more forgiving to user error and can give a reasonably accurate result.

Discussion Points:

- Which model type is being recommended?
 - It is a function of the process in which you are embedded. SCAA is preferred, but with lack of expertise a VPA approach is warranted.
- Which is preferred: generalized or custom built models?
 - Again it depends on the situation. The preferred option is a custom built model, but time and lack of expertise impede ability to always accomplish this result. In addition, the “black box” issue is an important one where lack of expertise can result in mis-use of the general model and the lack of necessary coding can result in the loss of coding skill and ability. Another problem is what to do when the generalized model does not do what you want and the code is so complicated or unavailable that the common user cannot alter it to fit the situation at hand.
 - A possible compromise might be a general model that provides a core that is highly accessible and available. Users could then take this core and alter the code in order to fit whatever situation is presented.

3 The selection of modelling approaches and software packages

3.1 Experience with software packages

It was important that the text on the software packages (presented in Section 2 above) reflected not just the impressions and conclusions of the model developers, who would inevitably have a different view on (say) ease of use, but also the wider group. Comments on issues raised in the ToRs were covered to a certain extent by the rapporteurs' summaries of the discussions, but the Chairs also wanted to conduct a more informal survey of impressions of software packages which did not incorporate the views of the package developers. In laying the groundwork for this, a form was passed around on which WKADSAM participants were asked to state, for each of the packages presented at the meeting, whether they were the developer, used the package regularly or occasionally, or had never used it. The results of this poll are given in Table 3.1.

Table 3.1. Results of a survey of software-package use among members of WKADSAM.

MODEL	NEVER USED	OCCASIONAL USE	REGULAR USE	DEVELOPER / CREATOR	TOTAL	% OCCASIONAL OR REGULAR USE
Adapt VPA	14	5	8	1	28	46%
ASAP	19	4	1	1	25	20%
B-ADAPT	18	6	1	1	26	27%
BREM	25	0	0	1	26	0%
CASAL	22	1	1	0	24	8%
CSA	21	2	1	1	25	12%
MULTIFAN-CL	22	0	2	0	24	8%
SAM	22	2	0	1	25	8%
SS3	18	3	2	1	24	21%
SURBA	18	3	2	1	24	21%
TINSS	23	0	0	1	24	0%
XSA	16	1	8	1	26	35%

There are some anomalies in this table: there were 32 participants in the meeting, and the Total column suggests that the survey was far from complete. However, the implication is clear. Of those who responded, the percentage that had used each package occasionally or regularly varied between 0% and 46% (with an emphasis on the standard VPA workhorses of ICES and NOAA). The mean percentage over all packages shown here was 17% - in other words, few of the participants of WKADSAM had much experience with the packages presented. It was therefore decided that a further review of perceptions of software packages would not be productive.

This means that the comments relating to the ToRs for each package presented in Section 2 do **not** represent a complete survey of the worldwide assessment community, but rather the views of the model developer and those WKADSAM participants who were motivated to speak up in plenary.

3.2 Case studies of model change: northern and southern hake

For several years until 2009 the northern and southern hake stocks were assessed in ICES using age-structured models: XSA for northern hake and a Bayesian statistical catch-at-age model for southern hake. However, tagging data experiments indicated that growth was considerably faster than what would be consistent with the age-length keys (obtained from otolith readings) used and no new otolith reading method was found. As a consequence, a decision was made to use only length-structured data in the assessments, which required abandoning the previously used assessment models.

A first decision to be taken was whether to develop a specific length-based assessment model for the hake stocks or to use one of the existing general models/packages, such as CASAL, GADGET, Multifan-CL or Stock Synthesis 3 (hereafter, SS). The choice was to use the general models/packages given the limited time available to do the work and the fact that they should have been extensively tested and were, hence, felt to be less prone to errors than purpose built models and accompanying code. This was felt to be very important given that the results of the hake assessments were to be used for the provision of advice.

The southern hake assessment coordinator had previous experience with GADGET (Globally applicable Area Disaggregated General Ecosystem Toolbox), which facilitated very much the process of setting up an assessment, so that was the option taken for this stock.

The choice was less clear for northern hake, with the scientists working in its assessment not having had any previous experience with any of the general models/packages. It was felt that CASAL, Multifan-CL or SS might be more suitable than GADGET, as they were specifically designed for assessment purposes, mainly for single-species assessments, whereas GADGET has been conceived as an ecosystem modelling tool (although it has also been used to conduct some stock assessments). Additionally, running times in GADGET are rather long, generally requiring over-night running, and getting uncertainty limits for the estimates, which was a desired feature in the assessment, appears to be difficult.

The available data for the northern hake assessment consisted of:

- Landings (tonnes) and length frequency distributions, quarterly from 1990 onwards and annually before 1990
- Quarterly discards estimates (tonnes) and length frequency distributions, but only for recent years, with many missing data in earlier years
- Relative abundance indices from surveys and corresponding length frequency distributions
- Growth information from mark-recapture experiments

The fact that there was no local expertise on SS, Multifan-CL or CASAL made the entire process considerably more difficult, starting with trying to decide which one to use, as it was not even clear whether the three of them could deal with the data available for the assessment. Multifan-CL was considered at first, but it appeared not to take in easily relative abundance indices, which are a crucial piece of information for this assessment. CASAL was very briefly considered but not enough information could be easily gathered (e.g. in terms of technical description of the model), so its use was not pursued. The choice was SS, which was understood to be capable of dealing with the types of data available. Coherent incorporation of the discards information was another important aim of the hake assessment, so a model that could cope

with the missing data were desired. SS could do that, internally estimating discards in the missing years. The fact that SS had been used for many groundfish stocks over many years also provided confidence that it would have been extensively tested in a wide range of relevant situations.

The above is not intended to imply that CASAL, GADGET or Multifan-CL could not have been set up appropriately to conduct this assessment, they might well have been, they merely provide a description of the thought process followed.

In terms of using the available data for northern hake, the drawbacks found with SS were:

- It was not possible to incorporate the available mark-recapture data, as SS seemed to require knowing the age of the fish at the time they were marked and released, which was unknown. The tagging data for hake consisted of the length of fish at capture and recapture times and the time elapsed between both events, essentially providing information on fish growth. The incorporation of this type of data as part of the likelihood in SS did not seem possible.
- Landings (tonnes) and corresponding length frequency distributions were available on a quarterly basis since 1990 but only on an annual basis in earlier years. SS could not handle the change in the time-step, so either the annual data had to be split between quarters in an ad-hoc manner or the assessment had to start in 1990 (the option taken).

Not being able to incorporate the mark-recapture data was seen as a drawback, but not serious enough to prevent the use of this model for the assessment, given that there was a lot of length-structured information from catch and surveys that could be incorporated and used as a source of information about growth. In terms of the time-step, if the model had been run on an annual rather than a quarterly step, a longer assessment period could have been considered. However, since length frequency distributions were to be used internally in order to estimate the growth parameter K (L_{∞} was fixed in advance), it was felt that a quarterly step would provide more accurate information on growth. Modal values could be followed through the quarters particularly well in the discards length frequency data and in some earlier surveys that were conducted quarterly.

A useful feature for a stock as complex as northern hake, fished in different areas and with a variety of gears, was that the commercial catch data (landings and discards) could be entered at fleet level. Seven fleets were defined, based on their main fishing zone and gear used. Landings (tonnes) by fleet are the only data that cannot be missing in SS, a value must be entered for every time-step in the assessment. For each fleet, discards (tonnes) and length frequency distributions of landings and discards separately were entered for the available time-steps. Selectivity of total catch and retention ogives are modelled at the fleet level and can either be assumed to be constant over time or allowed to vary in several ways. The parameters defining the selectivity curves and retention ogives can be estimated or fixed.

Learning how to set up a model in SS took considerable time and effort. This was, once again, not helped by the lack of local expertise. Setting up the northern hake model took about three months of two dedicated people, who were experienced with modelling but unfamiliar with SS at the outset.

In terms of assessment results it was found that whereas the estimated stock trends were quite stable across different model configurations (*i.e.* choices about natural

mortality, growth parameters, fleet selectivity-at-length curves, etc), all of them estimating an SSB increase and a decrease in F in recent years, the actual rates at which this increase and decrease were estimated to occur were very sensitive to the configuration used. The ICES benchmark workshop, which counted with the participation of the SS creator (Richard Methot), focused on analysing those sensitivities in detail and trying to find the most realistic model configuration. A main issue was the shape of the selectivity-at-length curves of the fishing fleets and surveys. If a flexible form was used for all of them, permitting them to be asymptotic or dome shaped for large lengths, the model always fitted selectivity curves that decreased to zero at large lengths and SSB estimates became very large. Imposing that selectivity curves be asymptotic for at least some of the fleets was felt to be more realistic, the idea being that if there were large fish in the population it was expected that at least some of the fleets would catch them. This idea was confirmed by the fact that earlier catch data, not included in the assessment, contained larger fish. A choice was made to impose asymptotic selectivity on the two fleets with the most persistent occurrence of large hake. Model choices regarding natural mortality and growth parameters were determined on the basis of model fit and sensitivity of results, pooling together information from the northern and southern hake stocks.

Weighting the different sources of information entering the likelihood was not straightforward and getting a weight configuration that was deemed appropriate required substantial input from the SS creator. The final choice of weights was essentially determined in an iterative way, trying to ensure that the values used matched the residual variability levels from the model fits.

The model set up with SS was accepted by the benchmark workshop as the new assessment model for the northern hake stock, replacing the previously used XSA. The benchmark workshop concluded that the new assessment was ready to be used for advice. Nonetheless, given the complete change of assessment methodology and the limited time available during the benchmark workshop, it was felt that some flexibility should be allowed in subsequent inter-benchmark years to continue improving the model configuration. Furthermore, some concern remained about the rates of increase and decrease estimated for SSB and F in recent years, which were considerably stronger than anticipated, casting some doubt on their realism. The ICES annual assessment working group that took place after the benchmark workshop echoed this concern and proposed that the assessment this year be accepted only as indicative of stock trends.

The current uncertainties in the assessment results for recent years are thought to be related to a lack of data information about the big individuals in the population, which are critical to determine SSB. No large individuals appear in the catch during the assessment period (starting in 1990), although they had been seen in earlier years. This is compounded by the fact that the surveys for this stock provide mostly information on young individuals and much less on the bigger ones. Hence, the model internally estimates SSB starting from recruitment (thought to be well estimated), which will then grow according to the estimated growth model, and subtracting individuals from the population according to the natural and fishing mortality rates estimates. Any errors in the growth or mortality rates (*e.g.* due to model misspecification, such as length- or time-varying parameters which are assumed to be constant, or potential biases in the catch data), will reduce the model's ability to estimate SSB and F accurately. Work at present focuses on trying to recover quarterly landings data from before 1990, when bigger individuals were present in the landings, so as to increase the contrast in the data. The possibility to standardize a cpue series from a

longline commercial fleet is also being explored, as this could provide an abundance index for the bigger individuals.

The southern hake assessment performed with GADGET was accepted at the benchmark workshop and has been subsequently used to provide catch advice for the stock.

The general conclusion drawn from the northern and southern hake experiences is that general models/packages such as CASAL, GADGET, Multifan-CL or SS can be effective assessment tools, particularly (but not exclusively) in situations where age data are not available but there is length information from catches and surveys, as well as stock abundance indices that can be used for tuning. In order to get robust recruitment and SSB estimates in these situations, the length spectrum represented in the catches and abundance indices must be as broad as possible, covering both the young and older fractions of the population. Learning to set up appropriate models using these general packages is not straightforward and it may be surmised that their use within ICES will remain quite limited unless local expertise is developed.

4 Conclusions

4.1 General recommendations

WKADSAM recognizes the importance of the distinction that ICES has made between benchmark and update stock assessments. The benchmark assessments should consider a wide range of modelling approaches and all available data. Conversely, update assessments should use a single model (denoted the *advisory model* here) for providing management advice. This advisory model should be robust to a wide range of uncertainty in the data and underlying processes. This feature should be demonstrated during the benchmark assessments and confirmed as much as possible during the update assessments using research models of different types. The use of management strategy evaluations (MSEs) to test the robustness of the advisory model is considered good practice.

There is a clear need for an advisory model for each stock. Once the advisory model has been determined through robustness testing, each annual update assessment should only require a review of the data, diagnostics, and output through an audit process. A sophisticated review of the advisory model is not required during each update assessment: instead, this type of review is more appropriate during the benchmark assessment. When the settings for the advisory model are determined during the benchmark assessment, they should not be changed during the update assessments without compelling evidence that there is a need to do so. For example, a benchmark setting may estimate a parameter that becomes inestimable in the update assessment due to age truncation in the data, requiring a change from the benchmark formulation during the update. The purpose of the advisory model is not to understand every underlying real-world process but to provide robust advice. Research models can be used to explore the underlying processes and lead the way to improved understanding during the next benchmark assessment.

WKADSAM considered both the types of stocks that are assessed in ICES and elsewhere, as well as the types of models that need to be developed for future stock assessments. Both sets were split into three types and ranked in order of importance. The order of importance of stocks for consideration is:

- 1) Stocks that are currently assessed incorrectly, e.g. forcing data into inappropriate framework. These should be reconsidered as soon as possible.
- 2) Stocks that are not currently assessed need to be assessed, if there is a customer requirement for such assessment.
- 3) Stocks that are currently assessed but could potentially be improved through research model improvement.

While Working Groups have been quite innovative in terms of conforming data into structures needed for particular software packages, the use of software packages that utilize the data in their original form is considered better practice. The large number of stocks that are currently not assessed poses a logistic challenge for Working Groups. Examination of these stocks for commonalities that could use similar modelling approaches could facilitate the completion of these assessments. Even stocks that do not currently have any diagnostic issues should continue to be examined through research models to ensure the advisory model remains robust to changing signals in the data.

The order of importance for model development and application is:

- 1) Modelling approaches that address stocks that are not currently assessed, e.g. data- and information-poor situations.
- 2) Modelling approaches that address stocks that could be assessed more appropriately, e.g. avoiding the need to slice length data for use in age-based models.
- 3) Modelling approaches that address stocks that are currently adequate but could be improved, e.g. spatial models.

Since there are already available a number of software packages for use with standard age-based assessments, the development of new approaches should focus on situations where these models cannot be directly applied due to lack of production ageing or understanding of basic biological characteristics such as growth and maturity. These situations could be approached from either a single modelling approach or by building case-specific models. All three modelling approaches should continue to receive attention in model development. While obviously related to the list above regarding stocks, this list is focused on the areas that ICES should support in terms of model development, either through specific training in workshops or explicit terms of reference for working groups.

WKADSAM encourages new members of ICES assessment WGs who have stock assessment responsibility to have the ability to write a simple stock assessment program (e.g. a production model) at a minimum. This would benefit the assessments and reviews in a number of ways. For the assessment scientist, it increases understanding of issues when using more advanced software packages and improves the ability to deal with unusual situations. Both of these benefits are the result of having worked with the coding of a stock assessment model, as opposed to just clicking options on a graphic user interface. The ICES community as a whole also benefits by increasing the future supply of stock assessment model creators, who will have to deal with situations and types of data that are not even considered currently. WKADSAM also recognizes the importance of including participants in Working Groups with specialized knowledge, e.g. survey history or regulations, which contribute to the final assessment, and so does not recommend this coding background for all participants.

In order to capture the important features of the system as parsimoniously and robustly as possible, selection of both the modelling approach and software package should be based on a thoughtful consideration of the available data, biology of the stock, management requirements, statistical principles, and (importantly) available expertise. The data, biology, and management issues require case-specific knowledge for each stock assessment. There are many statistical approaches, e.g. maximum likelihood and Bayesian, which can be applied to any given stock assessment; however care must be taken to do so in the most appropriate manner possible. The available network of experts either within the Working Group or to readily consult ensures that whichever modelling approach or software package is utilized that it is done so correctly. The highly flexible software packages such as Stock Synthesis, Multifan-CL, and CASAL can all be given data inputs that are nonsensical but still produce plausible (and wrong) output. Experts in such software packages can prevent this from happening.

The ultimate goal is to provide the best science advice to the management system. In order to do so, all of these features must be considered. For example, application of XSA when age data are not directly available is not good practice. WKADSAM recognizes there are two extreme approaches that could be taken to follow this advice:

all stocks use one model or all stocks use different models (which may or may not be bespoke). Neither is considered the best way to approach stock assessment. While use of one model for all stocks would improve standardization of outputs and understanding of that particular model; the use of different models for each assessment would allow complete customization of the stock assessment to each situation. The advantages of standardization should be recognized and used when appropriate. There should always be room for innovation in the research models for stock assessment, in order to advance the state-of-the-art in fisheries modelling and encourage modellers in the field.

4.2 Recommendations for the 2012 Conference

WKADSAM concluded by addressing ToR e), namely: "Prepare the groundwork for a following workshop in 2011 or 2012." The following broad conclusions were reached, which could be viewed as forming the basis for ToRs for the Conference itself:

- There is a clear need for the steering committee to be convened and to get planning immediately. Skip McKinnell of PICES anticipated that the steering committee would involve representatives from all the participating organizations, not just ICES, so these people will need to be identified and invited.
- There was a desire to hold a workshop with case studies prior to SISAM (the report of this workshop would be a big focus of at least one session in SISAM). This would consist of case studies of a representative sample of ICES stocks (maybe 10) from the full range of the data-availability spectrum – the aim would then be for experts in some of these different assessment approaches to compare their models with the standard ICES approaches to see what could be learned from changing to a new system.
- The workshop would therefore use real data instead of simulated data, and the majority of case studies would focus on data-limited situations (e.g. not just traditional age-based assessments).
- The workshop on case studies could keep momentum from this meeting through 2011 to the symposium in 2012.
- Some WKADSAM participants also thought that combining a large symposium with a smaller workshop immediately afterward would be productive - possible topic for follow-on workshop would be the "model of the future", with the focus on modelling approaches instead of software packages
- Potential talks at the symposium/conference could include:
 - Summaries by Regional Fisheries Management Organisations (RFMOs) of the models they use and why.
 - Papers on recent methodological advances in stock assessment, including (but not limited to):
 - Ecosystem approach and/or climate change (limited to 1–2).
 - Incorporating new types of data into assessments (e.g. physical oceanography; limited to 1–2).
 - Education and where is the field going (limited to 1).
 - Reports from the workshop described above.

- One important point is that these approaches need not be limited to the traditional single-area, single-species approach, but could (and should) be much more inclusive.

5 References

- Beare, D. J., Needle, C. L., Burns, F., and Reid, D. G. 2005. Using survey data independently from commercial data in stock assessment: An example using haddock in ICES Division VIa, *ICES Journal of Marine Science*, 62: 996–1005.
- Beare, D., Needle, C. L., Burns, F., Reid, D., and Simmonds, J. 2002. Making the most of research vessel data in stock assessments: examples from ICES Division VIa. ICES CM 2002/J:01.
- Cadrin, S. X. 2000. Evaluating two assessment methods for Gulf of Maine northern shrimp based on simulations. *Journal of Northwest Atlantic Fishery Science*, 27: 119–132.
- Collie, J.S., Kruse, G.H. 1998. Estimating king crab (*Paralithodes camtschaticus*) abundance from commercial catch and research survey data. In: Jamieson, G.S., Campbell, A. (Eds.), Proceedings of the North Pacific Symposium on Invertebrate Stock Assessment and Management. *Can. Spec. Publ. Fish. Aquat. Sci.*, 125: 73–83.
- Collie, J.S., Sissenwine, M.P. 1983. Estimating population size from relative abundance data measured with error. *Can. J. Fish. Aquat. Sci.*, 40: 1871–1879.
- Conser, R.J. 1995. A modified DeLury modelling framework for data-limited assessments : bridging the gap between surplus production models and age-structured models. Working document to the ICES Working Group on Methods of Fish Stock Assessment, Copenhagen, February 1995, 85 pp.
- Cook, R. M. 1997. Stock trends in six North Sea stocks as revealed by an analysis of research vessel surveys, *ICES Journal of Marine Science*, 54: 924–933.
- Cook, R. M. 2004. Estimation of the age-specific rate of natural mortality for Shetland sandeels, *ICES Journal of Marine Science*, 61: 159–164.
- Cotter, C., Fryer, R., Needle, C. L., Skagen, D., Spedicato, M.-T., and Trenkel, V. 2007b. A review of fishery-independent assessment models, and initial evaluation based on simulated data. Working Document for the ICES Working Group on Methods of Fish Stock Assessments Woods Hole, March 2007. Edited by Mesnil, B.
- Cotter, J., Fryer, R., Mesnil, B., Needle, C. L., Skagen, D., Spedicato, M.-T., and Trenkel, V. 2007c. A review of fishery-independent assessment models, and initial evaluation based on simulated data. ICES CM 2007/O:04.
- Deriso, R. B., Quinn II, T. J., and Neal, P. R. 1985. Catch-age analysis with auxiliary information, *Canadian Journal of Fisheries and Aquatic Sciences*, 42: 815–824.
- EFIMAS. 2007. Operational evaluation tools for fisheries management options. URL: <http://www.efimas.org/>. FLR Team (2006). FLR: Fisheries modelling in R. Version 1.2.1. Initial design by L. T. Kell and P. Grosjean.
- Fournier, D., and Archibald, C. P. 1982. A general theory for analysing catch at age data. *Canadian Journal of Fisheries and Aquatic Sciences*, 39: 1195–1207.
- Fryer, R. 2001. TSA: is it the way? Working document for the ICES Working Group on Methods of Fish Stock Assessment.
- Gudmundsson, G. 1986. Statistical considerations in the analysis of catch-at-age observations, *Journal du Conseil International pour l'Exploration de la Mer* 43: 83–90.
- Gudmundsson, G. 1987. Time series models of fishing mortality rates. ICES C.M. D:6.
- Gudmundsson, G. 1994. Time series analysis of catch-at-age observations. *Appl. Statist.* 43:117–126.

- Hilborn, R., and Mangel, M. 1997. The ecological detective: confronting models with data. Princeton Univ Pr.
- ICES. 2004a. Report of the Working Group on Methods of Fish Stock Assessments. ICES CM 2004/D:03.
- ICES (2006b). Report of the Working Group on Methods of Fish Stock Assessment, Galway, June 2006. ICES CM 2006/RMC:07.
- ICES. 2008a. Report of the Working Group on the Assessment of Demersal Stocks in the North Sea and Skagerrak. ICES CM 2008/ACOM:09.
- ICES. 2008b. Report of the Working Group on the Assessment of Northern Shelf Demersal Stocks. ICES CM 2008/ACOM:08.
- ICES-WGNSDS (2002). Report of the Working Group on the Assessment of Northern Shelf Demersal Stocks. ICES CM 2003/ACFM:04.
- Johnson, S. J., and Quinn II, T. J. 1987. Length frequency analysis of sablefish in the Gulf of Alaska, *Technical Report UAJ-SFS-8714*, University of Alaska, School of Fisheries and Science, Juneau, Alaska. Contract report to Auke Bay National Laboratory.
- MacCall, A. 2009. Depletion-corrected average catch: a simple formula for estimating sustainable yields in data-poor situations. *ICES Journal of Marine Science: Journal du Conseil*, 66(10):2267.
- Martell, S. J. D., Pine, W. E., and Walters, C. J. 2008. Parameterizing age-structured models from a fisheries management perspective. *Can. J. Fish. Aquat. Sci.*, 65:1586–1600.
- Maunder, M. M., Lee, H. H., Piner, K. R., and Methot, R. D. Estimating natural mortality within a stock assessment model: an evaluation using simulation analysis based on twelve stock assessments. Workshop on estimating natural mortality in stock assessment applications Seattle, WA, August 11–13, 2009.
- Mesnil B. 2005. Sensitivity of, and bias in, Catch-Survey Analysis (CSA) estimates of stock abundance. Kruse, G.H., V.F. Gallucci, D.E. Hay, R.I. Perry, R.M. Peterman, T.C. Shirley, P.D. Spencer, B. Wilson, and D. Woodby (eds.), *Fisheries assessment and management in data-limited situations*. Alaska Sea Grant College Program, University of Alaska Fairbanks, AK-SG-05-02: 757–782.
- Mesnil, B. 2003. The Catch-Survey Analysis (CSA) method of fish stock assessment: an evaluation using simulated data. *Fish. Res.*, 63: 193–212.
- Mesnil, B., Cotter, A. J. R., Fryer, R. J., Needle, C. L., and Trenkel, V. M. 2008. A review of fishery independent assessment models, and initial evaluation based on simulated data, *Aquatic Living Resources* 0000(0000): 0000–0000.
- Needle, C. L. 2004a. Absolute abundance estimates and other developments in SURBA. Working Document to the ICES Working Group on Methods of Fish Stock Assessment, IPIMAR, Lisbon 10–18 Feb 2004.
- Needle, C. L. 2004b. Data simulation and testing of XSA, SURBA and TSA. Working Paper to the ICES Working Group on the Assessment of Demersal Stocks in the North Sea and Skagerrak, Bergen, September 2004.
- Needle, C. L. 2004c. Testing TSA with simulated data. Working Paper to the ICES Working Group on the Assessment of Northern Shelf Demersal Stocks, Copenhagen, April 2004.
- Patterson, K. R., and Melvin, G. D. 1996. Integrated Catch At Age Analysis Version 1:2, *Scottish Fisheries Research Report*. FRS: Aberdeen.
- Pomarede, M., Simmonds, J., Hillary, R., McAllister, M., Kell, L., and Needle, C. L. 2006. Evaluating the management implications of different types of errors and biases in fisheries resources surveys using a simulation-testing framework. ICES CM 2006/I:28.

- Pope, J. G., and Shepherd, J. G. 1982. A simple method for the consistent interpretation of catch-at-age data, *Journal du Conseil International pour l'Exploration de la Mer* 40: 176–184.
- Quinn II, T. J., and Deriso, R. B. 1999. *Quantitative Fish Dynamics*, Oxford University Press, Oxford.
- Schirripa, M. J., Goodyear, C. P., and Methot, R. M., 2009. Testing different methods of incorporating climate data into the assessment of US West Coast sablefish. *ICES J. Mar. Sci.*, 66: 1605–1613.
- Yin, Y., and Sampson, D.B. 2004. Bias and precision of estimates from an age-structured stock assessment program in relation to stock and data characteristics. *North American Journal of Fisheries Management*, 24:865–879.

Annex 1: List of participants

	NAME	ADDRESS	TELEPHONE/TELEFAX	E-MAIL
1	Anders Nielsen	DTU Aqua - National Institute of Aquatic Resources Jægersborg Allé 1 DK-2920 Charlottenlund Denmark	+45 33963375	an@aqua.dtu.dk
2	Anthony Thompson	Northwest Atlantic Fisheries Organization PO Box 638 B2Y 3Y9 Dartmouth NS Canada	+1 902 468 7542	athompson@nafo.int
3	Benoit Mesnil	Ifremer Nantes Centre Rue de l'île d'Yeu PO Box 21105 F-44311 Nantes Cédex 03 France	+33 240 374009 +33 240 374075	benoit.mesnil@ifremer.fr
4	Brian Healey	Fisheries and Oceans Canada Northwest Atlantic Fisheries Center 80 East White Hills Road PO Box 5667 A1C 5X1 St John's NL Canada		Brian.Healey@dfo-mpo.gc.ca
5	Carmen Fernandez	Instituto Español de Oceanografía Centro Oceanográfico de Vigo Cabo Estay - Canido PO Box 1552 E-36200 Vigo (Pontevedra) Spain	+34 986 492111 +34 986 498626	carmen.fernandez@vi.ieo.es
6	Chris Darby	Centre for Environment, Fisheries and Aquaculture Science Lowestoft Laboratory Pakefield Road NR33 0HT Lowestoft Suffolk UK	+44 1502 524329 +44 7909 885 157 +44 1502 513865	chris.darby@cefas.co.uk
7	Christopher Legault [Chair]	National Marine Fisheries Services Northeast Fisheries Science Center 166 Water Street 2543 Woods Hole MA United States	+1 508 4952025 +1 508 4952393	chris.legault@noaa.gov

	NAME	ADDRESS	TELEPHONE/TELEFAX	E-MAIL
8	Coby Needle [Chair]	Marine Scotland Marine Laboratory Aberdeen PO Box 101 AB11 9DB Aberdeen UK	+44 1224 295456 +44 1224 295511	c.needle@marlab.ac.uk
9	Dan Goethel	University of Massachusetts Dartmouth 285 Old Westport Road 02747-2300 North Dartmouth MA United States		dgoethel@umassd.edu
10	Dana Hanselman	National Marine Fisheries Services AFSC Auke Bay Laboratories Ted Stevens Marine Research Institute PO Box 21668 99801 Juneau, AK United States		Dana.Hanselman@noaa.gov
11	Doug Butterworth	University of Cape Town Dept of Mathematics & Applied Mathematics 7701 Rondebosch South Africa	21 650 2343	doug.butterworth@uct.ac.za
12	Jan Horbowy	Sea Fisheries Institute in Gdynia ul. Kollataja 1 PL-81-332 Gdynia Poland	+48 58 735 6267 +48 58 7356 110	horbowy@mir.gdynia.pl
13	José De Oliveira	Centre for Environment, Fisheries and Aquaculture Science Lowestoft Laboratory Pakefield Road NR33 0HT Lowestoft Suffolk UK	+44 1502 527 7 27 +44 1502 524 511	jose.deoliveira@cefas.co.uk
14	Kalei Shotwell	National Marine Fisheries Services AFSC Auke Bay Laboratories Ted Stevens Marine Research Institute PO Box 21668 99801 Juneau, AK United States		Kalei.Shotwell@noaa.gov
15	Kurtis Trzcinski	Fisheries and Oceans Canada Bedford Institute of Oceanography 1 Challenger Drive PO Box 1006 B2Y 4A2 Dartmouth NS Canada		Kurtis.Trzcinski@dfo-mpo.gc.ca

	NAME	ADDRESS	TELEPHONE/TELEFAX	E-MAIL
16	Lionel Pawlowski	Ifremer Lorient Station 8, rue François Toullec 56100 Lorient France	+33 2 97 87 38 46 Fax +33 2 97 87 38 36	lionel.pawlowski@ifremer.fr
17	Matt Dunn	National Institute of Water and Atmospheric Research Wellington PO Box 14901 Wellington New Zealand		m.dunn@niwa.co.nz
18	Melissa Haltuch	National Marine Fisheries Services Northwest Fisheries Science Center 2725 Montlake Boulevard East 98112-2097 Seattle WA United States		Melissa.Haltuch@noaa.gov
19	Noel Cadigan	Fisheries and Oceans Canada Northwest Atlantic Fisheries Center 80 East White Hills Road PO Box 5667 A1C 5X1 St John's NL Canada	+1 709 772 5028 +1 709 772 4188	noel.cadigan@dfo-mpo.gc.ca
20	Norio Yamashita	Hokkaido National Fisheries Research Institute, Fisheries Research Agency Subarctic Fisheries Resources Division 116 Katsurakoi 085-0802 Hokkaido Japan	+81 154 92 1715 +81 154 91 9355	noriyama@affrc.go.jp
21	Ray Conser	University of California Center for Stock Assessment Research and Department of Applied Mathematics and Statistics (MS E-2) 1156 High Street CA 94607 Santa Cruz United States		ray.conser@noaa.gov
22	Richard D. Methot	National Marine Fisheries Services Northwest Fisheries Science Center 2725 Montlake Boulevard East 98112-2097 Seattle WA United States	+1 206 860-3365	richard.methot@noaa.gov
23	Ross Claytor	Fisheries and Oceans Canada 200 Kent Street K1A 0E6 Ottawa ON Canada		ClaytorR@mar.dfo-mpo.gc.ca

	NAME	ADDRESS	TELEPHONE/TELEFAX	E-MAIL
24	Shelton Harley	Secretariat of the Pacific Community B.P.D5 98848 Noumea Cedex New Caledonia		SheltonH@spc.int
25	Skip McKinnell	Institute of Ocean Sciences 9860 West Saanich Road PO Box 6 Sidney BC Canada		mckinnell@pices.int
26	Steve Martell	University of British Columbia UBC Fisheries Centre 6270 University Boulevard V6T 1Z4 Vancouver BC Canada		s.martell@fisheries.ubc.ca
27	Tetsuichiro Funamoto	Hokkaido National Fisheries Research Institute, Fisheries Reseach Agency Subarctic Fisheries Resources Division 116 Katsurakoi 085-0802 Hokkaido Japan	+81 154 92 1714 +81 154 91 9355	tetsuf@fra.affrc.go.jp
28	Timothy Earl	Centre for Environment, Fisheries and Aquaculture Science Pakefield Road NR33 0HT Lowestoft Suffolk UK	+44 (0) 1502 521303	timothy.earl@cefas.co.uk
29	Verena Trenkel	Ifremer Nantes Centre Rue de l'île d'Yeu PO Box 21105 F-44311 Nantes Cédex 03 France	+33 240 374157	verena.trenkel@ifremer.fr
30	Yuho Yamashita	Hokkaido National Fisheries Research Institute, Fisheries Reseach Agency Subarctic Fisheries Resources Division 116 Katsurakoi 085-0802 Hokkaido Japan	+81 154 92 1714 +81 154 91 9355	yuho@affrc.go.jp
31	Yukio Takeuchi	National Research Institute of Far Seas Fisheries, FRA 7-1, 5-chome Orido 424-8633 Shimizu-ku Shizuoka Japan		yukiot@fra.affrc.go.jp

	NAME	ADDRESS	TELEPHONE/TELEFAX	E-MAIL
32	Yuri A. Kovalev	Knipovich Polar Research Institute of Marine Fisheries and Oceanography 6 Knipovitch Street RU-183763 Murmansk Russian Federation	+7 8152 472 469 +7 8152 473 331	kovalev@pinro.ru

Note: Rebecca A. Rademeyer (University of Cape Town) was unable to attend the meeting, but subsequently contributed text to Section 2.14.1.

Annex 2: Software package descriptions

Annex 2.1: Adapt VPA

Model & Version	Adapt VPA 3.0.3
Category	(1) Age-based
Model Type	This version of virtual population analysis is part of the NOAA Fisheries Toolbox (NFT), but traces its lineage from Gavaris and Conser, and incorporates features introduced by Mohn, Powers, Restrepo, and Darby. As with all VPA models, it performs best in situations of high fishing mortality rates and strong production ageing programs that minimize uncertainty in the catch-at-age data. This model has been programmed to work with the NOAA Fisheries Toolbox (NFT) population simulator (PopSim) and has outputs that can be used in the NFT age structured projection program (AgePro).
Data used	Catch at age is required for all years and is entered as a single year by age matrix. Tuning indices (either from surveys or catch per unit of effort) are typically entered as age-specific time-series, but grouped ages can also be used. Weight at age is entered as three separate year by age matrices for catch, Jan-1 population biomass, and spawning-stock biomass. Biological parameters are entered as year by age matrices for natural mortality and maturity. Gaps are allowed in the tuning indices.
Model assumptions	Catch at age is assumed to have negligible error relative to the error in the tuning indices. Selectivity is derived from the fishing mortality values calculated back through each cohort. The model can be run with or without a plus group. The model assumes only a single area and one sex, so migration and sexual dimorphism are not explicitly modelled. There are no priors used in the model.
Estimated parameters	The parameters of the model are a set of population abundances at age in the year following the last year of catch data. The index catchability coefficients are nuisance parameters calculated internally from the observed and predicted values during each iteration. Optionally catch multipliers can be estimated, similar to Darby's B-Adapt approach.
Objective function	The objective function is simply the sum of squared residuals (optionally weighted) between the logarithms of the observed and predicted indices. Each index observation can be weighted independently. There are no priors.
Minimisation	The IMSL implementation of the Levenburg-Marquardt algorithm is used for minimization.
Variance estimates and uncertainty	Variances are available directly as a result of the Levenburg-Marquardt minimization as well as through optional bootstrapping of index residuals.
Other issues	The model has a built in retrospective analysis which successively removes years of data from the most recent year backwards, re-runs the model, collects the results, and provides graphical displays. Each retrospective "peel" can be opened independently in the GUI for full analysis if desired.
Quality control	Quite a bit of testing has been conducted; however, results are not easily accessible. Through use as the main age-based stock assessment model in the Northeast Fisheries Science Center for many years, it has been compared to many other models and always produced similar results when formulated similarly.
Restrictions	Single area model. No length information can be included directly (must be converted to age first).
Program language	The executable is written in Fortran with a GUI available for Windows machines.

Availability	<p>The program is available as an executable only. It is available with a GUI from the NOAA Fisheries Toolbox (NFT) website http://nft.nefsc.noaa.gov.</p>
References	<p>A reference manual is distributed with the GUI.</p> <p>Collie, J. S. 1988 . Evaluation of virtual population analysis tuning procedures as applied to Atlantic bluefin tuna. Collect. Vol. Sci. Pap ICCAT, 28: 203–220.</p> <p>Conser, R. J. and J.E. Powers. 1989. Extensions of the ADAPT VPA tuning method designed to facilitate assessment work on tuna and swordfish stocks. ICCAT Working Doc. scrs/89/43. 15pp.</p> <p>Gavaris, S. 1988. An Adaptive framework for the estimation of population size. CAFSAC Res. Doc. 88/29. 12pp.</p> <p>Gavaris, S. 1993. Analytical estimates of reliability for the projected yield from commercial fisheries. Can Spec Pub Fish Aquat Sci 120: 185–191.</p> <p>Gulland, J. A. 1965. Estimation of mortality rates. Annex. to Arctic Fisheries Working Group Report. ICES CM 1965, doc. No 3. 9pp.</p> <p>Mohn, R.K. and R. Cook. 1993. Introduction to sequential population analysis. NAFO Sci. Coun. Studies #17 110pp.</p> <p>Parrack, M. L. 1986. A method of analysing catches and abundance indices from a fishery. ICCAT Coll. Vol. Sci. Papers 24: 209–221.</p> <p>Patterson K.R. and G.P. Kirkwood. 1995. Comparative performance of Adapt and Laurec-Shepherd methods for estimating fish population parameters and in stock management. ICES Journal of Marine Science 52 (2): 183–196.</p> <p>Pope J. G. 1972. An investigation of the accuracy of virtual population analysis using cohort analysis. ICNAF Res. Bull. 9:65–74.</p>
Applications	<p>This model was been the workhorse of the Northeast Fisheries Science Center in the US. It has been used for most of the groundfish species in the region for the past two decades.</p>

Annex 2.2: AMAK

Model & Version	Assessment model from Alaska (AMAK)
Category	Age-based and can be used for data-poor
Model Type	Works well for single area, single-sex, multiple fisheries and indices allowed, time varying non-parametric recruitment implementation allows VPA-like to full separable fits to fishery catch-at-age data, flexible implementation of stock recruitment relationships (e.g. curve can be set for a “window” of years or not used at all). Mean weights-at-age input for all gear types. Continuous Baranof-equation for catch equation, season-specific indices, and allows sparse data (e.g. not available in all years),
Data used	Total catch tuned for solving Fs, catch-at-age for relative year-class strengths. Mean weights-at-age input for all fisheries and indices. Continuous Baranof-equation for catch equation, season-specific indices, and allows sparse data (e.g. not available in all years). Plus-group tracked.
Model assumptions	Typically penalized non-parametric selectivity (though parametric options available) for both fishery and indices. Uniquely, catchability for indices can be specified to apply to any range of valid ages. This feature allows catchability to be for specified ages which can be important if age-specific availability variability is acknowledged. Priors for key parameters optional (e.g. stock recruitment, natural mortality, catchability for indices, selectivity variability).
Estimated parameters	Depends on configuration but typically recruitment in each model year, annual components of fishing mortality, catchability for each index, catchability power coefficients relating index values to model predictions, age-specific component of fishing mortality (selectivity parameters), selectivity-at-age parameters for indices, stock recruitment parameters, Fmsy, Bmsy
Objective function	Penalized maximum-likelihood components of log-posterior distribution depends on configuration but typically consists of lognormal distribution on total catch biomass for each fishery (with annual input-specified uncertainty in “observed” totals), lognormal distribution for indices (also with annual input specified uncertainty in “observed” index values), multinomial distribution for composition data for fishery and indices when available (also with annual input specified uncertainty in “observed” composition sample sizes). Priors include lognormal distribution on recruitment values (following stock recruitment curve and variability estimates), lognormal penalties on selectivity variability, and lognormal distributions on natural mortality and index catchability.
Minimisation	ADMB quasi Newton
Variance estimates and uncertainty	Optionally asymptotic approximation to the joint posterior distribution, Hessian standard errors, or MCMC integration.
Other issues	
Quality control	Compared with earlier versions of stock synthesis and provided the same results (for the same model configurations). Extended to multispecies trophic interactions (Kinzey and Punt 2009).
Restrictions	No GUI
Program language	ADMB/C++
Availability	Yes, via e-mail (Jim.Janelli@noaa.gov) or older version at http://nft.nefsc.noaa.gov/

References	<p>Barbeaux, S., J.N. Ianelli, S. Gaichas, and M. Wilkins. 2008. Aleutian Islands walleye pollock SAFE. In Stock Assessment and Evaluation Report for the Groundfish Resources of the Bering Sea/Aleutian Islands Regions. North Pacific Fisheries Management Council, PO Box 103136, Anchorage, Alaska, 99510.</p> <p>Courtney, D.L., J. N. Ianelli, D. Hanselman, and J. Heifetz. 2007. Extending statistical age-structured assessment approaches to Gulf of Alaska rockfish (<i>Sebastes</i> spp.). In: Heifetz, J., DiCosimo J., Gharrett, A.J., Love, M.S, O'Connell, V.M, and Stanley, R.D. (eds.). Biology, Assessment, and Management of North Pacific Rockfishes. Alaska Sea Grant, University of Alaska Fairbanks. pp 429–449.</p> <p>Hanselman, D. H., J. Fujioka, C. Lunsford, and C. Rodgveller. 2009. Alaskan Sablefish. In Stock assessment and fishery evaluation report for the groundfish resources of the GOA and BS/AI as projected for 2010. North Pacific Fishery Management Council, 605 W 4th Ave, Suite 306 Anchorage, AK 99501. pp. 353–464.</p> <p>Hanselman, D. H., S. K. Shotwell, J. Heifetz, J. Fujioka, and J. N. Ianelli. 2009. Gulf of Alaska Pacific ocean perch. In Stock assessment and fishery evaluation report for the groundfish resources of the Gulf of Alaska as projected for 2008. North Pacific Fishery Management Council, 605 W 4th Ave, Suite 306 Anchorage, AK 99501. Pp. 743–816.</p> <p>Ianelli, J.N., S. Barbeaux, T. Honkalehto, S. Kotwicki, K. Aydin, and N. Williamson. Assessment of the walleye pollock stock in the Eastern Bering Sea. In Stock Assessment and Evaluation Report for the Groundfish Resources of the Bering Sea/Aleutian Islands Regions. North Pacific Fisheries Management Council, PO Box 103136, Anchorage, Alaska, 99510.</p> <p>Lowe, S., J. Ianelli, M. Wilkins, K. Aydin, R. Lauth, and I. Spies. 2009. Stock assessment of Aleutian Islands Atka mackerel. In Stock Assessment and Evaluation Report for the Groundfish Resources of the Bering Sea/Aleutian Islands Regions. North Pacific Fisheries Management Council, PO Box 103136, Anchorage, Alaska, 99510.</p> <p>Shotwell, S.K., D. Hanselman, and D. Clausen. 2009. Gulf of Alaska rougheye rockfish. In Stock assessment and fishery evaluation report for the groundfish resources of the Gulf of Alaska as projected for 2010. North Pacific Fishery Management Council, 605 W 4th Ave, Suite 306 Anchorage, AK 99501. pp. 993–1066.</p>
Applications	<p>Aleutian Islands Pollock</p> <p>Bering Sea and Aleutian Islands Atka mackerel</p> <p>Alaska sablefish</p> <p>Gulf of Alaska rockfish</p> <p>Others...</p>

Annex 2.3: ANP

Model & Version	ANP
Category	(1) Age-based;
Model Type	This is a statistical catch-at-age model in R, using splines to model selectivity at age, and discarding at age. The shape of the splines can vary over time. The model was specifically designed to deal with those stocks where discards estimates are available for some years, but not in others. Because it “reconstructs” the discards estimates internally, it does need fishery-independent tuning indices in the age range of the discards.
Data used	Landings and discards at age, tuning indices. The discards can have gaps in the time-series. For those years that discards need to be reconstructed, fisheries independent tuning indices should be available. Currently, no plus group calculations are done inside the model.
Model assumptions	Constant catchability in the surveys, but not in the fleets. The overall level of fishing mortality is estimated by year, the selectivity pattern over ages is a smooth function of time. Discarding at age can either be fixed in time, or a smooth function of time. No plus group assumptions are made. The maturity, growth and natural mortality are not estimated inside the model, and so are assumed to be without error. The model assumes a single stock, so without a spatial component and without migration. No sexes are included in the model, but it could be run for both sexes separately. The starting values for the optimizer are chosen at random. In order to ensure independence of the outcome from the starting values, a large number of runs are done from different random starting values
Estimated parameters	Catchability at age for tuning indices, landings, and discards, landings and discards at age, stock numbers-at-age and fishing mortality-at-age, all including uncertainty estimates.
Objective function	The objective function is the combined log likelihood of the model discards, landings, and tuning series (log-likelihood). No ad-hoc weighting occurs of these sources.
Minimisation	R optimizer (optim).
Variance estimates and uncertainty	Resampling from multivariate normal distribution using inverse numerically derived hessian.
Other issues	The model is written in R and can easily adopted to change assumptions.
Quality control	The model has been published in the ICES JMS. In its application model selection criteria can be used to test alternative hypotheses. Source code is open
Restrictions	If used to estimate discards, no estimates can be given outside the time span for which fisheries independent tuning indices are available. Also, at sufficient number of years with discards information (probably approximately for one full cohort) should be available to be able to reconstruct discards.
Program language	R
Availability	The program is available as source code through the authors
References	Aarts and Poos 2009 ICES Journal of Marine Science 66: 763–771: “Comprehensive discard reconstruction and abundance estimation using flexible selectivity functions”
Applications	North Sea plaice, exploratory assessment since 2009

Annex 2.4: ASAP

Model & Version	ASAP 2.0.20
Category	(1) Age-based
Model Type	The Age Structured Assessment Program (ASAP) is an age-structured model that uses forward computations assuming separability of fishing mortality into year and age components to estimate population sizes given observed catches, catch-at-age proportions, and indices of abundance. Discards can be treated explicitly. It is a relatively simple model with a limited number of options, making it well suited for use as an introductory statistical catch-at-age model. It has been programmed to work with the NOAA Fisheries Toolbox (NFT) population simulator (PopSim) and has outputs that can be used in the NFT age structured projection program (AgePro).
Data used	The data used are catch time-series by fleet, catch-at-age proportions by fleet, tuning indices (either from surveys or fishery catch per unit of effort) which can be either age specific (East Coast of US style) or a total index with age proportions (West Coast of US style), weight-at-age (different matrices allowed for catch, Jan-1 population, and spawning stock), and the basic biological parameters of year by age matrices for natural mortality and maturity. The oldest age is always a plus group. Gaps are allowed in catch-at-age proportions and index time-series.
Model assumptions	Selectivity can be estimated by age directly or through either a single logistic or double logistic equation for each time block. Catchability can be either a single value for the time-series or else allowed to vary according to a random walk. Fishing mortality follows the Baranov catch equation. The oldest age is a plus group with no structure within the bin. The model is only for one area, so migration terms are allowed. The model is not separated by sex, so sexual dimorphism cannot explicitly be modelled. Priors are allowed on a number of parameters and assume a lognormal distribution, creating a penalized likelihood as the objective function.
Estimated parameters	Fleet and time block specific selectivity can be modelled as age specific parameters, or as a single logistic or double logistic function. Fleet specific fishing mortality multipliers (fully selected) by year. Index catchability, possibly by year if random walk. Index selectivity by age, single logistic, or double logistic. Beverton–Holt stock–recruitment relationship. Recruitment deviations from expected curve. Population abundance-at-age in first year.

Objective function	<p>Lognormal error distribution assumed for</p> <p>Total catch in weight</p> <p>Total discards in weight</p> <p>Indices</p> <p>Stock recruitment deviations</p> <p>Multinomial distribution assumed for</p> <p>Catch at age</p> <p>Discards at age</p> <p>Index proportions at age</p> <p>Optional penalties which assume lognormal error distribution</p> <p>Two stock recruitment parameters (relative to initial guesses)</p> <p>F in year 1 by fleet (relative to initial guesses)</p> <p>Changes in F from year to year</p> <p>Catchability in year 1 for each index (relative to initial guesses)</p> <p>Catchability random walk</p> <p>Initial population age structure (relative to initial guess)</p> <p>Weights can be applied to each component of the objective function.</p>
Minimisation	Standard AD Model Builder minimization using phases.
Variance estimates and uncertainty	Standard AD Model Builder Hessian variance estimates are provided and MCMC can be conducted as well.
Other issues	The model has a built in retrospective analysis which successively removes years of data from the most recent year backwards, re-runs the model, collects the results, and provides graphical displays. Each retrospective “peel” can be opened independently in the GUI for full analysis if desired.
Quality control	Quite a bit of testing has been conducted; however, results are not easily accessible. Particular emphasis on comparison between ASAP and VPA for data with strong retrospective patterns demonstrated both models performed similarly.
Restrictions	Single area model. No length information can be included directly (must be converted to age first).
Program language	AD Model Builder for executable, GUI on Windows machines.
Availability	Full source code provided in Technical Manual distributed with GUI. Can be downloaded from the NOAA Fisheries Toolbox (NFT) website http://nft.nefsc.noaa.gov .
References	<p>User Manual and Technical Manual both distributed with GUI.</p> <p>Legault, C.M. and V.R. Restrepo. 1999. A flexible forward age-structured assessment program. Int. Comm. Cons. Atl. Tunas, Coll. Vol. Sci. Pap. 49(2): 246–253.</p>
Applications	ASAP has been used as an assessment tool for Atlantic herring (NEFSC), Atlantic mackerel (NEFSC), ICES horse mackerel, red grouper (SEFSC), yellowtail flounder (NEFSC), Pacific sardine (SWFSC), Pacific mackerel (SWFSC), Greenland halibut (ICES), Northern Gulf of St Lawrence cod (DFO), Gulf of Maine cod (NEFSC), Florida lobster (FFWCC), and fluke (NEFSC).

Annex 2.5: Bayes Discards

Not yet provided.

Annex 2.6: Bayesian State-Space version of delay-difference model

Model & Version	Bayesian State-Space version of delay-difference model.
Category	(2) length/stage based
Model Type	Implementation of Bayesian State-Space version of delay-difference model given in Meyer and Millar (1999) with mean weight modification from Hilborn and Walters (1992). Two stages used; recruits and commercial size scallops. Model modified to include submodel for clappers (empty joined shells) as a proxy measure for natural mortality because of occurrence of catastrophic mortality events. Model used to estimate current and past biomass, exploitation and forecast one year ahead for TAC advice.
Data used	Catches, survey biomass estimates for recruits and commercial size scallops, survey estimates of clappers, growth parameters based on survey data.
Model assumptions	Selectivity is assumed to be knife-edge and catchability to the survey is estimated in the model. Assumptions on priors (non-informative) and distributions for data.
Estimated parameters	Annual natural mortality, biomass, catchability to survey, exploitation all from posterior distribution.
Objective function	State-space formulation with process and observation errors.
Minimisation	Gibbs/Metropolis sampling
Variance estimates and uncertainty	Variances from posterior distribution.
Other issues	Model can be modified for selectivity patterns.
Quality control	Standard testing of sampling (e.g. Brooks-Gelman-Rubin method) plus model results evaluated using the posterior predictive distribution of the input data. Also evaluate projections from previous years.
Restrictions	Nothing to date.
Program language	WinBugs 1.4.3
Availability	Source code: See Smith and Lundy (2002) for early form of model and Jonsen et al. (2009) for more recent version.
References	<p>Deriso, R. B. 1980. Harvesting strategies and parameter estimation for an age-structured model. <i>Canadian Journal of Fisheries and Aquatic Sciences</i> 37: 268–282.</p> <p>Hilborn, R., and Walters, C. J. 1992. Quantitative fisheries stock assessment: Choice, dynamics and uncertainty. Chapman and Hall, New York.</p> <p>Jonsen, I.D., A. Glass, B. Hubley, and J. Sameoto. 2009. Georges Bank ‘a’ Scallop (<i>Placopecten magellanicus</i>) Framework Assessment: Data Inputs and Population Models. DFO Can. Sci. Advis. Sec. Res. Doc. 2009/034. iv + 76 p.</p> <p>Meyer, R., and Millar, R. B. 1999. Bayesian stock assessment using a state-space implementation of the delay difference model. <i>Canadian Journal of Fisheries and Aquatic Sciences</i> 56: 37–52.</p> <p>Smith, S. J., and Lundy, M. 2002b. Scallop production in Area 4 in the Bay of Fundy: Stock status and forecast. DFO Canadian Science Advisory Secretariat Research Document 2002/018: 90 p.</p> <p>Smith, S. J., Lundy, M., Sameoto, J., and Hubley, B. 2008. Scallop Production Areas in the Bay of Fundy: Stock Status for 2008 and Forecast for 2009. DFO Canadian Science Advisory Secretariat Research Document 2008/22: vi + 108 p.</p>
Applications	Bay of Fundy scallops since 2002; Georges Bank scallops since 2009.

Annex 2.7: BBM

Model & Version	BBM (Two-stage biomass-based model)
Category	(2) length/stage based
Model Type	<p>Bayesian state-space model with stochastic recruitment process and deterministic population dynamics. Model dynamics described in terms of biomass and separated into two stages (recruits -age 1- and older individuals). Biomass decrease due to growth and natural mortality encapsulated into a unique parameter (g) age and time invariant. Catches are just considered instantaneous removals from the available population. Observation equations consist on total biomass and age 1 biomass proportion from the research surveys.</p> <p>The model was constructed for short-lived species with highly variable recruitment, as an alternative to fully age-structured models. The model is based on the work by Roel and Butterworth (2000). Similar models are those in Collie and Sissenwine (1983), Mesnil (2003) and Trenkel (2008).</p>
Data used	Total biomass and age 1 biomass proportion from the research surveys are included in the observation equations. Total catch and age 1 catch (in mass) before and after the research surveys are accounted for as removals of the population. Intrinsic growth and natural mortality rates are assumed to be known and age and time invariant. Fractions of the years when the surveys and the catches occur are also needed.
Model assumptions	<p>The biomass decrease parameter (intrinsic growth and natural mortality) is age and time invariant. The catchability of total biomass from the research surveys are assumed to be constant in the whole time-series. The age 1 proportion from the research surveys are assumed to be unbiased estimates of the age 1 proportion in the population. This implicitly means that the surveys's catchability is constant across ages.</p> <p>The prior distributions are centred at values that are considered realistic and chosen to have substantial but not unreasonably large dispersion. Sensitivity to the prior distributions of recruitment and initial biomass is tested in Ibaibarriaga et al. (2008).</p>
Estimated parameters	Initial biomass, average and precision (inverse of variance) of the normal process error for log-recruitment, survey catchability for total biomass from the research surveys and precisions of the observation equations of total biomass and age 1 biomass proportion from the research surveys. In addition, the biomass decrease rate due to growth and natural mortality can also be estimated.
Objective function	The joint posterior probability density function (pdf) of the unknowns is the product of the pdf's of observations, states and priors. The total biomass from the research surveys is lognormally distributed. The age 1 proportion from the research surveys follows a beta distribution. The stochastic recruitment process is lognormal. The prior distributions for the survey catchabilities, the initial biomass, the mean of the recruitment process, the variance-related parameter of the age 1 proportion observation equations and the biomass decrease parameter are lognormal, whereas the prior distributions of the precisions of the total biomass observation equations and the precision of the recruitment process are gamma distributed. In addition, an indicator function (takes value 1 or 0) that indicates whether the restrictions imposed by the catches (biomass must be larger than the catches) are fulfilled or not is included.
Minimisation	Bayesian inference conducted using Markov chain Monte Carlo (MCMC) techniques.
Variance estimates and uncertainty	The joint posterior distribution of the parameters is obtained from the MCMC runs.

Other issues	Currently the model is specifically designed for the Bay of Biscay anchovy, but it could be modified and adopted to different stocks and assumptions. The model is currently being extended. The new model allows separating the natural mortality and growth process and splitting them by age group (recruits and olders) and incorporates total catch and age 1 catch proportion into the observation equations.
Quality control	The model has been tested on simulated datasets that were generated conditioned on the model itself. No robustness test has been performed.
Restrictions	Due to the high correlations between the parameters when all the parameters are estimated the problem is undetermined and the solutions might be affected by the chosen prior distributions. The model does not provide fishing mortality estimates; instead harvest rates (catch/biomass) are used. Also, as the model is biomass-based, recruitment refers to age 1 biomass at the beginning of the year.
Program language	The model is written in WinBUGS and it is run from R using the R2WinBUGS library. Analysis of the results is conducted in R using the coda library.
Availability	The program is available from the authors on request.
References	Ibaibarriaga, L., Fernández, C., Uriarte, A., Roel, B.A. (2008). A two-stage biomass dynamic model for Bay of Biscay anchovy: a Bayesian approach. ICES Journal of Marine Science 65: 191–205. Anchovy assessment working groups from 2005 onwards (WGMHSA 2005–2007, WGANC 2008, WGANSA 2009–2010) and benchmark workshop on short-lived species (WKSHORT 2009)
Applications	From 2005 onwards it is used to assess the Bay of Biscay anchovy stock (Latest reference WGANSA 2010). Within the EU-project SARDONE it has also been applied to the Aegean Sea anchovy.

Annex 2.8: BREM

Model & Version	Two-stage biomass random effects model (BREM)
Category	(2) stage based: recruits and total population
Model Type	<p>Two stage biomass model: $B_t = R_t + g_{t-1} B_{t-1}$</p> <p>$B_t$ = total population biomass</p> <p>R_t = the recruitment in biomass in year t</p> <p>g_{t-1} = biomass growth rate during year $t-1$</p> <p>Both recruitment R_t and biomass growth g_t are treated as random effects:</p> <p>$\log(R_t) \sim N(\mu_R, \sigma_R^2)$</p> <p>$\log(g_t) = \log(g_{t-1}) + \varepsilon_t$ with $\varepsilon_t \sim N(-0.5\sigma_g^2, \sigma_g^2)$</p> <p>Assumptions: effects of catches on the interannual variation of the integrative parameter g_t are either random or, if not, sufficiently small not to matter.</p> <p>Application context: Situations with only survey indices and no commercial catches available, no age data but information for recruits and total population. Recruits don't have to be age 0.</p>
Data used	<p>The observation model has two components: total biomass b_t at time t (recruits included) and recruits r_t.</p> <p>$\log(b_t) \sim N(\log(q_b B_t), \sigma_{I2})$</p> <p>$\log(r_t) \sim N(\log(q_r R_t), \alpha \sigma_{I2})$</p> <p>Both indices are assumed to follow lognormal distributions with the variance for the recruit index being a multiple of that for the total biomass index and each have separate constants of proportionality.</p> <p>This formulation has been developed for the case where one survey method is used to obtain a total biomass index and a different method for the recruit index although the recruit index might rely partly on the same information.</p> <p>Data gaps: no problem apart from in first and final year</p>
Model assumptions	<p>To make the model identifiable, catchability q_b is set to 1 but q_r is estimated. Hence all population biomass estimates are relative to the index for which q is fixed to 1. $\alpha = 1$, but any other value could be chosen for a given case study. If two survey series are used for the same time period, separate constants of proportionality are fitted for each survey series, but again constraining the constant of proportionality of one of the survey series to one.</p> <p>Instead of estimating survey observation error σ_1 the coefficient of variations CVI is estimated.</p>
Estimated parameters	<p>Parameters that can be estimated by the model:</p> <p>$\theta = \{\mu_R, \sigma_R, g_1, \sigma_g, q_r, CVI, B_1\}$</p>
Objective function	<p>Estimation of model parameters θ is carried out by maximum likelihood based on the observation vector $y = (b_1, \dots, b_n, r_1, \dots, r_n)$ which has conditional density $f_\theta(y u)$ where $u = (R_2, \dots, R_n, g_2, \dots, g_n)$ is the vector of latent random variables with marginal density $h(u)$. The marginal likelihood function is obtained by integrating out u from the joint density</p> $f_\theta(y u)h_\theta(u)$ $L(\theta) = \int f_\theta(y u)h_\theta(u) du$ <p>The joint penalized loglikelihood is $PL(\theta) = \log(f_\theta(y u)) + \log(h_\theta(u))$.</p>
Minimisation	AD model builder
Variance estimates and uncertainty	Parameter estimation by Maximum likelihood; variance from fitted Hessian.

Other issues	<ul style="list-style-type: none"> - Different model version exist, for example with fixed total mortality across years or use of several survey series (explained in Trenkel 2008) - Parameter estimation is sensitive to suitable starting values; otherwise there can be convergence problems or a crash;
Quality control	<p>Testing by simulation</p> <p>Which parameters are identifiable (Trenkel 2008)</p> <p>Performance for simulated data (Mesnil et al 2009)</p>
Restrictions	<p>temporal changes in catchability cannot be handled</p> <p>indices for recruits and total stock should not be correlated</p>
Program language	AD-model builder
Availability	Code and executable available from author
References	<p>Mesnil, B., Cotter, A. J. R., Fryer, R. J., Needle, C. L., Trenkel, V. M. 2009. A review of fishery-independent assessment models, and initial evaluation based on simulated data. <i>ALR</i>, 22: 207–216.</p> <p>Trenkel, V. M. 2008. A two-stage biomass random effects model for stock assessment without catches: What can be estimated using only biomass survey indices? <i>CJFAS</i>, 65: 1024–1035.</p> <p>Trenkel, V. 2009. Anchovy assessment in the Bay of Biscay using a two-stage biomass random effects model (BREM). Working document to ICES Benchmark Workshop on Short-lived Species, 31 August - 4 September 2009, 5 pp.</p>
Applications	Application to anchovy in the Bay of Biscay; working document presented to ICES Benchmark Workshop on Short-lived Species (Trenkel 2009).

Annex 2.9: CSA

	Catch-Survey Analysis (CSA)
Model & Version	April 2005 (BM) + October 2008 Version 3.1.1 (NFT)
Category	(2) stage based \pm (3) data-poor (no age data)
Model Type	<p>Collie & Sissenwine (1983) two-stage model: estimates time-series of recruitment and stock size (in number) given t-s of total catches, and survey (or cpue) indices in number for recruits and post-recruits (a super plus-group).</p> <p>Designed to assess stocks where age determination is impossible or very uncertain (e.g. crustaceans, hake, etc.) but a recruit stage can be distinguished from all larger/older fish. Superior to surplus-production in that it does account for change in dynamics due to variation in recruitment.</p>
Data used	<p>Total catches (in number); "survey" indices in number for recruits and post-recruits; natural mortality M (possibly varying by year); timing of catch in year.</p> <p>Mean weights by stage, only needed to convert final estimates of stock numbers to biomass.</p> <p>Relative weight of measurement errors in survey indices on recruits (both versions) and of process errors (mixed-error version) wrt measurement errors on post-recruits indices.</p> <p>Very sporadic missing data can be handled (only in observation-error version): estimates for the corresponding years are unreliable but the impact is short-lived.</p>
Model assumptions	<p>Only considers catchability in survey (not in fishery). Fully recruited catchability assumed constant through time-series (a variant in R allows for one step change). Catchability of recruits is a user-set fraction of that of fully recruited.</p> <p>Recruitment must be such that all so-called recruits move to the post-recruit stage in the following time-step (year), i.e. cannot be a sum of 1–2–3 "real" age groups.</p> <p>Populations assumed closed (no migration).</p>
Estimated parameters	<p>Observation-error (CSAo): Stock size of recruits in all years except the last; stock size of fully recruited in first year; fully recruited catchability (computed as GM), i.e. Y+1 parameters if Y years of data.</p> <p>Mixed-error (CSAme & NFT): Recruits indices in all years but the last; fully recruited indices in all years; fully recruited catchability, i.e. 2Y parameters if Y years of data.</p> <p>Optional in both: the catchability ratio between recruits and post-recruits can be sought by grid search ("SSQ profiling").</p>
Objective function	<p>Observation-error: Weighted sum (user defined weights) of 2 sums of squared log-residuals between model-predicted and observed survey indices, one for recruits, the other for fully recruited.</p> <p>Mixed-error: Weighted sum (user defined weights) of 3 sums of squared log-residuals between model-predicted and observed survey indices, one for recruits, one for fully recruited, and one for process error terms. The process error component can be assumed additive or multiplicative.</p>
Minimisation	Minimisation by non-linear least squares, using a Marquardt-Levenberg algorithm (same in R and Fortran implementations). The jacobian can be provided explicitly, otherwise derivatives are done numerically.
Variance estimates and uncertainty	<p>Approximate CV of parameters based on the Hessian.</p> <p>Retrospective analysis.</p> <p>Non-parametric model-conditioned bootstrap (Fortran port only): residuals from base run drawn randomly, for each error source independently, and added to fitted indices. Table of percentiles produced for biomasses and q; no bias correction (yet) on percentiles.</p>

Other issues	<p>Recruitment in the terminal year is not estimated; it can be inferred from the terminal survey index (knowing q and s), or taken as a recent average (like in some VPA assessments).</p> <p>Three- or four-stage extensions have been developed to assess crab stocks in the North Pacific, where CSA is used routinely.</p> <p>If catch is known by stage, then harvest rates can be estimated by dividing by the estimated stock sizes, and carried to catch forecasts similar to the conventional age-based procedure.</p>
Quality control	<p>Most testing was about sensitivity rather than robustness to violations in model assumptions ("clean" artificial data). Tests showed effect of errors in setting the catchability ratio (can be huge), the natural mortality, the timing of catch, the relative weights on errors, depletion rate and trajectory in the population, wrong allocation to recruits, and the effect of trends or step changes in catchability; several of these effects are qualitatively similar to those on VPA.</p>
Restrictions	<p>Does not handle multiple surveys per stage.</p> <p>Trials on simulated and real data indicate that absolute estimates of stock size are mostly sensitive to the assumed ratio of recruits catchability (s) which has to be set by users based on external information and analyses, but trends are less sensitive to s.</p> <p>Estimates weakly sensitive to weights of error sources in objective function.</p> <p>Estimated q negatively correlated with assumed M and s.</p> <p>No effect of the selection profile in the survey or the fishery.</p> <p>Performance degrade, even for trends, when indices too noisy.</p> <p>Unaccounted trends in survey q result in biased stock size estimates, but retrospective plots do not detect this.</p> <p>Like VPA, CSA works best when the (implied) fishing mortality is high.</p>
Program language	<p>A "French" version maintained as R scripts, with older implementations in Fortran for Unix and Windows. The R version require installing and loading the library minpack.lm from CRAN.</p> <p>The NFT version is an installable program with graphic user interface for use on Windows machines. The core program is written in Fortran.</p>
Availability	<p>The NFT version is accessible through the NFT home page at: http://nft.nefsc.noaa.gov/CSA.html</p> <p>R scripts, Fortran sources and (depending on your Internet firewall) executables for the French version can be obtained from B. Mesnil at Ifremer (Benoit.Mesnil@ifremer.fr); they are also available in the software library at ICES.</p>

References	<p>Origin: Collie, J.S. and Sissenwine, M.P. (1983). Estimating population size from relative abundance data measured with error. <i>Canadian Journal of Fisheries and Aquatic Sciences</i>, 40:1871–1879.</p> <p>NFT and French versions coded from:</p> <p>Conser, R.J. (1994). Stock assessment methods designed to support fishery management decisions in data-limited environments: development and application. PhD thesis. School of Fisheries, University of Washington, Seattle, 292pp. and Conser, R. J. (1995). A modified DeLury modelling framework for data-limited assessments: bridging the gap between surplus production models and age-structured models. Work. Doc. to the ICES WG on Methods of Fish Stock Assessment, Copenhagen, February 1995, 85 pp.</p> <p>Sensitivity tests: Mesnil, B. (2003). The Catch-Survey Analysis (CSA) method of fish stock assessment: an evaluation using simulated data. <i>Fisheries Research</i>, 63: 193–212.</p> <p>Mesnil, B. (2004). A crash test of Catch-Survey Analysis (CSA) using the NRC simulated data. Working Document to the ICES Working Group on Methods of Fish Stock Assessment, Lisbon, 2004.</p> <p>Mesnil, B. (2005). Sensitivity of, and bias in, Catch-Survey Analysis (CSA) estimates of stock abundance. <i>Fisheries Assessment and Management in Data-Limited Situations</i>. G. H. Kruse, V. F. Galluci, D. E. Hay <i>et al.</i>, Alaska Sea Grant College Program, University of Alaska Fairbanks. AK-SG-05-02: 757–782.</p>
Applications	<p>CSA is routinely used in the USA for the assessment of crustaceans (shrimp, king crab, blue crab) in advisory settings (see Cadrin, Collie, Helser, Murphy, Zheng and co-authors).</p> <p>First exploratory trials in ICES area were on Nephrops in the Bay of Biscay but fell on the difficulty of separating recruits in (annual) length compositions.</p> <p>In 2005 WGNSSK used CSA to cross-check the VPA results for the problematic whiting assessment. At that period HAWG tried CSA for the sprat assessment. WGDEEP and WGNEW also considered using CSA but the outcome is unknown.</p>

Annex 2.10: MULTIFAN-CL

Model & Version	MULTIFAN-CL, version 1. (But we have not yet implemented a structured versioning system – this is currently being done)
Category	<p>(1) Age-based; (2) length/stage based; (3) data-poor</p> <p>Age structured (i.e. population at age is modelled), but length-based (uses length and weight data to inform age and some processed (selectivity) have a length-based option.</p>
Model Type	<p>MULTIFAN-CL is a computer program that implements a statistical, length-based, age-structured model for use in fisheries stock assessment. The model is a convergence of two previous approaches. The original MULTIFAN model (Fournier et al. 1990) provided a method of analysing time-series of length–frequency data using statistical theory to provide estimates of von Bertalanffy growth parameters and the proportions-at-age in the length–frequency data. The model and associated software were developed as an analytical tool for fisheries in which large-scale age sampling of catches was infeasible or not cost-effective, but where length–frequency sampling data were available. MULTIFAN provided a statistically based, robust method of length–frequency analysis that was an alternative to several ad hoc methods being promoted in the 1980s. However, MULTIFAN fell short of being a stock assessment method as the endpoint of the analysis was usually estimates of catch-at-age (although later versions included the estimation of total mortality and yield-per-recruit).</p> <p>The second model (actually the first, in terms of chronology) was that introduced by Fournier and Archibald (1982). The FA model was a statistical, age-structured model in which estimates of recruitment, population-at-age, fishing mortality, natural mortality and other estimates useful for stock assessment could be obtained from total catch and effort data and catch-at-age samples. In principle, the estimates of catch-at-age obtained from the MULTIFAN model could be used as input data to the FA model and a complete stock assessment analysis conducted.</p> <p>Such a sequential approach to length-based stock assessment modelling had several serious limitations. First, it was extremely unwieldy. Second, it was difficult to represent and preserve the error structure of the actual observed data in such a sequential analysis. This made estimation of confidence intervals for the parameters of interest and choice of an appropriate model structure for the analysis problematic. It was clear that an integrated approach was required, one that modelled the age-structured dynamics of the stock, but which recognized explicitly that the information on catch-at-age originated with length–frequency samples.</p> <p>The early versions of MULTIFAN-CL, which were developed for an analysis of South Pacific albacore (Fournier et al 1998), provided the first attempt at developing a statistical, length-based, age-structured model for use in stock assessment. Subsequent versions of the software have added new features, the most important of which have been the inclusion of spatial structure, fish movement and tagging data in the model (Hampton and Fournier 2001). MULTIFAN-CL is now used routinely for tuna stock assessments by the Oceanic Fisheries Programme (OFP) of the Secretariat of the Pacific Community (SPC) in the western and central Pacific Ocean (WCPO). Beginning in 2001, the software gained additional users, with stock assessment applications to North Pacific blue shark, Pacific blue marlin, Pacific bluefin tuna, North Pacific swordfish and Northwest Hawaiian lobster underway or planned.</p>

Data used	<p>Catch in number or in weight (but must be consistent within fishery). Missing data allowed if effort is available.</p> <p>Effort in consistent units within fishery. Missing data allowed if catch is available.</p> <p>Length frequency. Missing data allowed.</p> <p>Weight frequency. Missing data allowed.</p> <p>Tagging data, for whatever period may be covered by the programme.</p> <p>Minimum data requirements would be catch, either length or weight frequency data for each defined fishery, maybe possible to configure the model without effort data but not ideal.</p>
Model assumptions	<p>Selectivity may be estimated as a length or age based process. A number of methods used to constrain parameterization, including functional forms and cubic splines.</p> <p>Catchability may be specified as constant over time, or varying via a random walk process. Deviations in the latter constrained by prior of mean zero and specified variance. Flexible time-stepping for random walk. If assumed constant over time, catchability may be linked across fisheries of the same gear in different model regions to allow cpue to indicate relative abundance spatially. Seasonality may be estimated as a separate process. A separate random effect called effort deviations is also modelled and constrained by priors of mean zero and specified variance.</p> <p>Fishing mortality is the product of selectivity, catchability and effort.</p> <p>Growth may be estimated in VB or Richards formulations. Also allow deviations from the growth curve for a specified number of age classes.</p> <p>Natural mortality may be estimated as an age invariant or age-specific parameter set. Smoothing penalties used to constrain variability.</p> <p>Recruitment may be specified as occurring with monthly to annual periodicity. This specification defines an 'age class' and the time-stepping in the model. Estimate an overall recruitment scaling factor, with temporal and regional deviates, all of which may be constrained by priors.</p> <p>Movement among defined model regions occurs and parameters may be related to age in a simple functional form.</p> <p>Maturity at age is specified and used to define spawning biomass. There are currently no sex-specific aspects to the model.</p>
Estimated parameters	<p>The following parameters may be estimated:</p> <p>Growth</p> <p>Natural mortality</p> <p>Mean recruitment, spatial and temporal deviates. Mean recruitment may be integrated into a Beverton–Holt SRR and steepness may be estimated or specified.</p> <p>Selectivity</p> <p>Catchability mean, seasonality, temporal deviates, effort deviates.</p> <p>Movement</p> <p>Reporting rates if tagging data are used</p>
Objective function	<p>Components for catch (lognormal), length frequency (robust lognormal), weight frequency (robust lognormal), tagging (negative binomial with option for zero inflation). Priors for all estimated parameters, additional smoothing penalties to constrain variability and avoid overfitting. Option for exact catch, in which case there is no catch likelihood. Weighting for length and weight frequency specified as 'effective sample size' and may be fishery specific. Weighting for tag data controlled by specified or estimated overdispersion parameters for the negative binomial.</p>
Minimisation	<p>Automatic differentiation – same source code as ADMB.</p>

Variance estimates and uncertainty	Variance-covariance matrix for model parameters derived from Hessian. Variance and confidence intervals for dependent quantities may be derived using the Delta method. Probability distributions for certain management quantities, e.g. F/FMSY and B/BMSY are obtained by likelihood profiling. Structural and data uncertainty handled in grid-wise structural sensitivity analyses.
Other issues	Model is continually being extended. There is also a java based utility for examining results and an R library for generating various diagnostics and results summaries. A stock projection capability is incorporated, with option for stochastic projections incorporating variability in recruitment, estimated terminal population and projected effort deviations. Either catch or effort can be used to drive the projections. Software to facilitate set up of projections is currently under development.
Quality control	Ad hoc testing regime for checking new code. Several structured simulation testing studies, e.g. Labelle, M. 2005. Testing the MULTIFAN-CL assessment model using simulated tuna fisheries data. Fisheries Research 71, 311–334.
Restrictions	Of course there are many! Cannot currently handle sex-specific data or model multiple stocks.
Program language	C++
Availability	Executables may be downloaded from www.multifan-cl.org . Source code may be provided under certain circumstances.
References	See www.multifan-cl.org . Fournier, D., and Archibald, C.P. (1982). A general theory for analysing catch-at-age data. Canadian Journal of Fisheries and Aquatic Sciences 39, 1195–1207. Fournier, D.A., Sibert, J.R., Majkowski, J., and Hampton, J. (1990). MULTIFAN: a likelihood-based method for estimating growth parameters and age composition from multiple length frequency datasets illustrated using data for southern bluefin tuna (<i>Thunnus maccoyi</i>). Canadian Journal of Fisheries and Aquatic Sciences 47, 301–317. Fournier, D.A., Hampton, J., and Sibert, J.R. (1998). MULTIFAN-CL: a length-based, age-structured model for fisheries stock assessment, with application to South Pacific albacore, <i>Thunnus alalunga</i> . (pdf - 287k) Canadian Journal of Fisheries and Aquatic Sciences 55, 2105–2116. Hampton, J., and D.A. Fournier. (2001). A spatially disaggregated, length-based, age-structured population model of yellowfin tuna (<i>Thunnus albacares</i>) in the western and central Pacific Ocean. (pdf - 4502k) Marine and Freshwater Research 52, 937–963.
Applications	Too numerous to note here. Routinely used for annual assessments of skipjack, yellowfin and bigeye tuna in the western and central Pacific, and albacore in the South Pacific. Has also been used from time to time for North Pacific albacore, North Pacific bluefin, swordfish (North and Southwestern Pacific), blue marlin, striped marlin, blue shark (North Pacific), Hawaiian rock lobster, Indian Ocean yellowfin, Atlantic Ocean bigeye and Atlantic Ocean albacore.

Annex 2.11: Non-equilibrium production model using Z and trawl surveys data

Model & Version	Non-equilibrium production model using Z and trawl surveys data
Category	(3) data-poor
Model Type	Is a Biomass dynamic approach that use Z and Biomass indices. A series of data of estimates of Biomass indices and Total Mortality rates derived from trawl surveys are used for fitting a non-equilibrium production model. It allows a rough estimation of the value of fishing mortality rate that produces the Maximum Sustainable Yield (FMSY). Reliability of results is linked with the level of contrast in the time-series.
Data used	Couples of estimates of Z and of an index of biomass. Z can be a mean value of the last 2 or 3 years. An estimate of M is needed
Model assumptions	Exploitation pattern unchanged along the analysed time
Estimated parameters	Parameters r and K (an index) of the logistic growth model and FMSY
Objective function	
Minimisation	Two equivalent procedures: 1) minimization of the sum of the squared deviations between logarithms of observed and estimated values of Biomass by changing the seed values of r and K 2) minimization of the log likelihood value by changing the seed values of r and K.
Variance estimates and uncertainty	<p>The confidence bounds for K and r can be estimated through the construction of a likelihood profile (Venzon & Moolgavkar, 1988; Schnute, 1989). In order to obtain the likelihood profile, the likelihood function is defined and the maximum likelihood estimated. Assuming lognormal error, the equation that has to be minimized is the following:</p> $-LnL = (n-1)Ln\left[\sqrt{\frac{2\pi}{n}}\sum_{y=1}^n\left(LnI_y - Ln\hat{I}_y\right)^2\right] + \frac{n-1}{2}$ <p>The estimation of confidence bounds of the parameters is based on the observation that in this particular case error shows a distribution that can be approximated by the χ^2 distribution with m degrees of freedom (Punt & Hilborn, 1996). In consequence, within the confidence interval for K and r, (in this case at p=95% and 1 degree of freedom) there will be included all the values for which twice the difference between the negative of the log-likelihood and the negative of the log-likelihood corresponding to the maximum likelihood estimates is less than 3.841 (χ^2 1, 0.05).</p>
Other issues	Being the biomass estimates derived from trawl surveys only indices of the absolute biomass at sea, the approach does not allow the estimation of an absolute value for MSY nor of K but only the level of F that produces the MSY
Quality control	
Restrictions	Not useful when fishing pressure remains stable along the dataseries. Fishing pattern has to be remained unchanged along the study period
Program language	Excel spreadsheet
References	<p>Caddy J., Defeo O. 1996. Fitting the exponential and logistic surplus yield models with mortality data: some explorations and new perspectives. Fish. Res. 25:39–62</p> <p>Abella, A. 2007 Assessment of European hake with a variant of a non-equilibrium Biomass Dynamic Model using exclusively trawl surveys data. WG SAC GFCM-FAO Athens, September 2007.</p>

Applications	The model has been used in Stock Assessment Committee-GFCM_FAO meetings and in working groups of the Mediterranean Sub-group SGMED of STECF
--------------	---

Annex 2.12: SAD

Model & Version	SAD (Separable-ADAPT VPA, version: ICES 2009)
Category	Age-based
Model Type	<p>A linked separable VPA and ADAPT VPA model, so that different structural models are applied to the recent and historical periods. The separable component applies to the most recent period, while the ADAPT VPA component applies to the historical period. Model estimates from the separable period initiate a historical VPA for the cohorts in the first year of the separable period. Fishing mortality at the oldest true age (age 10) in the historical VPA is calculated as the average of the three preceding ages (7–9, ignoring the 1982 year class where applicable), multiplied by a scaling parameter that is estimated in the model. In order to model the directed fishing of the dominant 1982 year class, fishing mortality on this year class at age 10 in 1992 is estimated in the model. The scaling parameter deals with the directed fishing on this year class once it enters the plus-group. The model also incorporates potential fecundity per kg as a function of fish weight, and realized fecundity per kg to help scale the model.</p>
Data used	<p>Egg production estimates, used as relative indices of abundance and catch-at-age data (numbers). Weights-at-age in the stock and maturity-at-age vary temporally, but are assumed to be known without error. Natural mortality and the proportions of fishing and natural mortality before spawning are fixed and year-invariant. Fecundity data are potential fecundity vs. fish weight data for the years 1987, 1992, 1995, 1998, 2000 and 2001, and a realized fecundity 'prior' distribution for 1989, with a mean and CV derived from a normal distribution in log-space, which covers (with a 95% probability) the range of realized fecundity values reported by Abaunza et al. (2003).</p>
Model assumptions	<p>The separable period assumes constant selection-at-age, and requires estimation of fishing mortality age- and year-effects (the former reflecting selectivity-at-age) for ages 1–10 and the final x years for which catch data are available (x being the length of the separable period). Selectivity at age 8 is assumed to be equal to 1. The length of the separable period should be balanced against the precision of model estimates and whether there is any indication, from the log-catch residuals, that the separable assumption no longer holds.</p> <p>The fishing mortality-at-age 10 (the final true age) is equal to the average of the fishing mortalities at ages 7–9 (ignoring the 1982 year class where applicable) multiplied by a scaling parameter estimated within the model. The fishing mortality-at-age 10 in 1992 (applicable to the 1982 year class) is estimated separately. The plus-group fishing mortality is assumed equal to that of age 10.</p> <p>A dynamic plus group is assumed (plus group this year is the sum of last year's plus group and last year's oldest true age, both depleted by fishing and natural mortality). The plus group modelled in this manner allows the catch in the plus group to be estimated, and making the assumption that log-catches are normally distributed allows an additional component in the likelihood, fitting these estimated catches to the observed plus-group catch.</p>
Estimated parameters	<p>The parameters treated as "free" in the model (i.e. those estimated directly) are: (1) Fishing mortality year effects for the final four years for which catch data are available; (2) Fishing mortality age effects (selectivities) for ages 1–10 (except for selectivity at age 8 which is set to 1); (3) scaling parameter for fishing mortality-at-age 10 relative to the average for ages 7–9 (ignoring the 1982 year class where applicable); (4) fishing mortality on the 1982 year class at age 10 in 1992; (5) realized fecundity parameter, relating realized fecundity to potential fecundity, and therefore also relating estimated SSB to the egg production estimates; (6) potential fecundity parameters (intercept and slope), relating potential fecundity to fish weight.</p>

Objective function	The estimation is based on maximum likelihood. There are five components to the likelihood, corresponding to egg estimates, catches for the separable period, catches for the plus-group, potential fecundity vs. fish weight, and realized fecundity. The variance of each component is estimated, apart from that associated with realized fecundity for which a CV is input.
Minimisation	The minimization routine in ADMB is used.
Variance estimates and uncertainty	Estimates of precision may be calculated by the several methods available in ADMB, the simplest (based on the delta method) and quickest being the one used most often.
Other issues	The model is readily extendible to account for other sources of data (as was done for the fecundity data)
Quality control	A range of simulation tests, as described in De Oliveira et al. (2010) were performed.
Restrictions	Custom model to handle the particular feature of western horse mackerel.
Program language	ADMB
Availability	Source code freely available in ICES folders, and from jose.deoliveira@cefas.co.uk
References	<p>Abaunza, P., Gordo, L., Karlou-Riga, C., Murta, A., Eltink, A. T. G.W., Garcí'a Santamarí'a, M. T., Zimmermann, C., et al. 2003. Growth and reproduction of horse mackerel, <i>Trachurus trachurus</i> (Carangidae). <i>Reviews in Fish Biology and Fisheries</i>, 13: 27–61.</p> <p>De Oliveira, J. A. A., Darby, C. D., and Roel, B. A. 2010. A linked separable–ADAPT VPA assessment model for western horse mackerel (<i>Trachurus trachurus</i>), accounting for realized fecundity as a function of fish weight. – <i>ICES Journal of Marine Science</i>, 67: 916–930.</p> <p>ICES. 2009. Report of the Working Group on Widely Distributed Stocks (WGWIDE), 2–8 September 2009, Copenhagen, Denmark. ICES CM 2009/ACOM:12. 563 pp.</p>
Applications	Western horse mackerel (ICES) since 2000 in various versions. Current version since 2009.

Annex 2.13: SAM

Model & Version	State-space Assessment Model (SAM). 0.2-r
Category	(1) Age based
Model Type	SAM is a time-series model designed to be an alternative to the (semi) deterministic procedures (VPA, Adapt, XSA, ...) and the fully parametric statistical catch-at-age models (SCAA, SMS, ...). Compared to the deterministic procedures it solves the problem of falsely assuming catches-at-age are known without errors, and in addition the problem of selecting appropriate so-called 'shrinkage', and in certain cases convergence problems in the final years. Compared to fully parametric statistical catch-at-age models SAM avoids the problem of fishing mortality being restricted to a parametric structure (e.g. multiplicative), and many problems related to having too many model parameters compared to the number of observations (e.g. borderline identification problems, convergence issues, asymptotic results, ...)
Data used	Total catch-at-age data and survey indices. Natural mortality M (possibly varying by year). Mean weights by age in stock and catch. Proportion mature.
Model assumptions	Log catches and log indices are assumed to follow normal distributions. Fishing mortalities are assumed to follow random walks (separate for age groups). Natural mortality, proportion mature and weights are assumed known. Further the model is built around the usual stock and catch equations. This simple model further has the advantage that it can easily be adapted to cases where assumptions need to be adjusted.
Estimated parameters	Observation errors, process errors, survey-catchabilities, and depending on configurations the stock-recruitment parameters are estimated. In addition the fishing mortalities and stock sizes are predicted (also for the historical period).
Objective function	The joint likelihood of observations, unobserved random variables (fishing mortalities and stock sizes), and model parameters is set up, then the marginal likelihood is computed by integrating* out the unobserved random variables. The marginal likelihood is optimized to give the maximum likelihood estimates of the model parameters. *) Note: The integration is carried out via the highly efficient Laplace approximation built into AD Model Builder, but has been validated via an unscented Kalman filter, and importance sampling.
Minimisation	Quasi-Newton algorithm aided by automatic differentiation (as implemented in AD Model Builder)
Variance estimates and uncertainty	Based on Hessian and delta method. Profile likelihood and MCMC validation is available without additional coding.
Other issues	Many features are not covered above. The main advantage of having a very simple base model as described is that case specific issues can be dealt with easily. For instance for North Sea cod the model includes estimation of a catch multiplier which explains a mismatch between surveys and catches. For some stocks a technical creep has been tried, and for others climate variables has been included to improve the stock-recruitment relationship. Finally, the maximum likelihood framework of this model allows statistical significance tests to be performed.
Quality control	SAM has been tested via simulation studies, output diagnostics, and by comparing to results from other models.
Restrictions	At the time of writing SAM is only a single area and single species model. If the fishing mortality in the last year jumps by many times the levels seen in the past, then the time-series nature of the model will dampen the jump. This is equivalent to the effect of year-shrinkage, but in SAM it is objectively estimated.

Program language	AD Model Builder. To make it more accessible an online version is available for certain stocks, and that is based on a mix of php and R scripts calling the main program.
Availability	http://stockassessment.org
References	<p>Origin of state-space models in Assessment:</p> <p>Gudmundsson, G. (1987). Time series models of fishing mortality rates. ICES C.M. (d:6)</p> <p>Gudmundsson, G. (1994). Time series analysis of catch-at-age observations. Appl. Statist. (43):117–126.</p> <p>Fryer, R. (2001). TSA: is it the way? ICES Working document for the Working Group on Methods of Fish Stock Assessment.</p> <p>The Laplace approximation and its use in AD model Builder:</p> <p>H. Skaug and D. Fournier (2006). Automatic approximation of the marginal likelihood in non-gaussian hierarchical models. Computational Statistics & Data Analysis. (56):699–709.</p> <p>Detailed description of the model:</p> <p>Report of the Working Group on Methods of Fish Stock Assessment (WGMG 2009). Section 8.</p>
Applications	<p>SAM is currently run for the following stocks in ICES:</p> <p>Kattegat Cod, Western Baltic Cod, Sole in 3A, Eastern Baltic Cod, North Sea Sole, Plaice in 3A, and North Sea Cod.</p> <p>Of these the state-space assessment model is primary for the first three stocks, and included as exploratory for the remaining.</p> <p>In addition to the stocks mentioned above it has been applied to applied to other stocks (Western Baltic spring-spawning herring, North Sea Haddock, 3PS Cod, and Georges Bank Yellowtail Flounder) for testing purposes, and has performed well.</p>

Annex 2.14: Stock Synthesis (SS)

Model & Version	Stock Synthesis (SS) version 3.10b
Category	(1) Age-based; (truly both age- and length-based)
Model Type	<p>SS is a generalized age- and length-based model that is very flexible with regard to the types of data that may be included, the functional forms that are used for various biological processes, the level of complexity and number of parameters that may be estimated. Numbers at age for each yearclass are tracked for each of several cohorts defined in terms of sex, mean growth pattern, and birth season. The recruitment of each cohort can be apportioned among areas and movement among areas can occur seasonally. The distribution of size-at-age for each cohort follows a normal distribution to allow for implementation of length-selectivity and to derive fishery specific body weight-at-age. Further, each cohort can be subdivided by size among several morphs (platoons) in order to allow for fishery size-selectivity to cause size survivorship within each cohort.</p>
Data used	<p>There is no minimum data requirement. Gaps can be included in all data sources, although catch is normally modelled as known for each time-step. Data types include:</p> <ul style="list-style-type: none"> catch, discards (in biomass or as a fraction of landings), indices of abundance (surveys or fishery cpue), mean body weight (across sampled ages), length compositions, age compositions, weight compositions, conditional age-at-length compositions, mean length-at-age, mean weight-at-age, tag releases and recaptures, stock composition data (e.g. microchemistry or genetic data) among the model identities defined as growth patterns environmental data <p>Bins structure for composition data are separate from bins for population dynamics calculations and includes aggregation in largest and smallest bins.</p>
Model assumptions	<p>Numerous selectivity options are available as a function of length or age and age- and length-based selectivity can be combined. Fishing mortality can be applied as a continuous rate or in the middle of the season using Pope's approximation.</p> <p>Fleets and surveys can mirror selectivity of each-other or use different forms. Population plus group is aggregated, but the maximum number of ages is unrestricted.</p> <p>Maturity is logistic, growth follows von-Bertalanffy or Richards growth-curve. Natural mortality may be a single value, a piecewise linear function of age, a Lorenzen function, or a vector of values at each age (with or without interpolation across seasons).</p> <p>Movement can be included between any pairs of areas in spatial models and movement rates is a 2-parameter dog-leg shaped function of age.</p> <p>Recruitment is a single value in each year based on various spawner-recruit options, which is then assigned to areas, genders, growth-patterns, growth-morphs, etc. according to a set of parameters that may be fixed or time-varying.</p> <p>Annual total recruitment is defined as a lognormal deviation from a spawner-recruitment function, or from a constant mean value. Substantial controls are provided to account for the consequences of estimating recruitment variability in data-poor eras of the modelled time-series.</p>

Estimated parameters	<p>Long list of possibilities is difficult to fully enumerate. Possible estimated parameters include those controlling growth, weight-at-length, maturity, selectivity at length and/or age, spawner-recruit relationship, annual recruitment, distribution of recruitment among various partitions of population structure, movement rates, tagging mortality and reporting rates, catchability (including possible non-linear relationship with abundance), parameters controlling offsets in the above relationships across genders or growth patterns, and parameters controlling temporal variation in any other parameters.</p> <p>In general, all parameters may be fixed across all years or time-varying according to a block structure, a set of random deviations, a random walk, or a smooth trend over time.</p> <p>Priors can be included on any parameter as normal, lognormal, beta.</p>
Objective function	<p>Objective function is a combination of components for</p> <ul style="list-style-type: none"> cpue or abundance index (lognormal or normal) fishery Discard biomass (normal) fishery or survey Mean body weight (normal) fishery or survey Length composition (multinomial) fishery or survey age composition (multinomial) fishery or survey Mean size at age (normal) Initial equilibrium catch (normal) Recruitment deviations (lognormal) Random parameter time-series deviations (normal) Parameter priors Penalty on negative abundance
Minimisation	<p>Minimisation is implemented using standard ADMB process. Minimization occurs in phases, and all parameters may be assigned to a phase in which estimation will begin.</p>
Variance estimates and uncertainty	<p>Variance estimates for all estimated parameters and numerous derived quantities are calculated either the Hessian matrix or from MCMC calculations, both implemented using standard ADMB algorithms.</p> <p>Parametric bootstrap datasets can be generated in order to evaluate the reproducibility of model results.</p>
Other issues	<p>Any other features or issues not covered above, e.g. the possibilities for model extensions.</p>
Quality control	<p>Numerous tests have been conducted using this model. Those published in peer reviewed literature include Yin and Sampson (2004), which reached the conclusion that “For all the output variables examined the estimates appeared to be median-unbiased”, wq and Schirripa et al. (2009) which focused on incorporating climate data, but provided an additional check of the ability of the model to estimate parameters using simulated data.</p> <p>Various ongoing research projects have determined that SS is capable of estimating parameters used to simulate data. These include the work of Maunder et al. (2009) and separate projects being conducted by Ian Taylor, Tommy Garrison, and Chantel Wetzel, all associated with the University of Washington. The simulations studies have included data simulated within stock synthesis as well as data generated from independent operating models written in R.</p> <p>SS has been used for dozens of stock assessments around the world. The area of highest used is on the US Pacific Coast. Numerous stock assessments conducted by NMFS scientists at the Northwest and Southwest Fisheries Science centers using SS have been reviewed by a stock assessment review (STAR) panel which includes independent CIE reviewers. These assessments are then reviewed by the Scientific and Statistical Committee of the Pacific Fishery Management Council.</p>

Restrictions	Single species assessments only. Growth transition matrices (e.g. those used for invertebrates) are not possible. Recruitment is a function of global spawning output, so true metapopulation structures are not yet possible.
Program language	ADMB
Availability	<p>The model and a graphical user interface are available from the NOAA Fisheries Stock Assessment Toolbox website: http://nft.nefsc.noaa.gov/. Only executable code is routinely distributed, along with a manual and sample files. However, under certain circumstances, source code may be obtained from the author upon request and with agreement to certain restrictions.</p> <p>An set of R routines to process and view model output is available from http://code.google.com/p/r4ss/. These routines were initially developed by Ian Stewart and Ian Taylor.</p>
References	<p>Maunder, M.M., Lee, H.H. Piner, K.R. and Methot, R.D. Estimating natural mortality within a stock assessment model: an evaluation using simulation analysis based on twelve stock assessments. Workshop on estimating natural mortality in stock assessment applications Seattle, WA, August 11–13, 2009. (submitted to CJFAS)</p> <p>Methot, R. D. 1990. Synthesis model: an adaptable framework for analysis of diverse stock assessment data. <i>Int. North Pac. Fish. Comm. Bull.</i> 50, pp. 259–277.</p> <p>Methot, R. D. 2000. Technical Description of the Stock Synthesis Assessment Program. National Marine Fisheries Service, Seattle, WA. NOAA Technical Memorandum NMFS-NWFSC-43. 46 pp.</p> <p>Methot, R. D. 2009. Stock assessment: operational models in support of fisheries management. In <i>The Future of Fishery Science in North America</i>, pp. 137–165. Ed. by R. J. Beamish, and B. J. Rothschild. Fish and Fisheries Series, 31. 736 pp.</p> <p>Methot, R. D. 2010. User Manual for Stock Synthesis Model Version 3.10 Updated Feb 20, 2010.</p> <p>Methot, R.D. and Taylor, I.G., 2010. Modelling the variability of recruitment in fishery assessment models. In review.</p> <p>Schirripa, M. J., Goodyear, C. P., and Methot, R. M., 2009. Testing different methods of incorporating climate data into the assessment of US West Coast sablefish. <i>ICES J. Mar. Sci.</i> 66, 1605–1613.</p> <p>Sheng-PingWang, Mark N. Maunder, Alexandre Aires-da-Silva, and William H. Bayliff. 2009. Implications of model and data assumptions: An illustration including data for the Taiwanese longline fishery into the eastern Pacific Ocean bigeye tuna (<i>Thunnus obesus</i>) stock assessment. <i>Fisheries Research</i> 97: 118–126.</p> <p>Yin, Y. and Sampson, D.B. 2004. Bias and precision of estimates from an age-structured stock assessment program in relation to stock and data characteristics. <i>North American Journal of Fisheries Management</i> 24:865–879.</p>
Applications	SS has been used for dozens of stock assessments around the world. The area of highest used is on the US Pacific Coast where it was first applied in the late 1980s. Application species for production assessments have included dozens of groundfish stocks, numerous tuna stocks, other large and small pelagics, sufclams, toothfish, sharks and various other fish. Exploratory analyses have been conducted for shrimps and various other species.

Annex 2.15: SURBA

Model & Version	SURBA (version 3.0)
Category	Age-based.
Model Type	SURBA uses a separable mortality model to estimate total mortality and relative stock abundance from one or more age-structured or biomass survey indices. Survey catchabilities must be given as input parameters, and uncertainty estimation for mortality and recruitment is currently performed via the delta method. SURBA was intended to provide survey-based assessments for those stocks for which catch data are either unavailable or unreliable. It is widely used in ICES assessment working groups for exploratory data analysis, and is also used to provide final assessments for two stocks (as of 2009). It is best suited for use with surveys which are not very noisy, and for stocks with a good interannual contrast in recruitment. Three more recent versions (SURBAR, SURBA+ and SAS-SURBA) are in development. A module (FLSURBA) is available in the FLR library, but is not actively supported.
Data used	SURBA requires age-structured and (optionally) biomass survey indices, along with age-structured natural mortality, mean weights-at-age and maturity. Survey catchabilities and SSQ-weightings are determined by the user. In the current version, up to 30 age-structured and 10 biomass indices can be included, although there must always be at least one age-structured index. Missing data are treated as such in the SSQ minimization.
Model assumptions	Survey catchability and SSQ weightings must both be provided by the user or assumed to be equal across all ages and years (although see the SURBA+ implementation by Noel Cadigan, DFO). Natural mortality is also provided externally and simply subtracted from estimated total mortality to generate fishing mortality (so the method is actually estimating total mortality). Mortality is assumed to be separable into age and year components; the estimated age component is then applied to all years. All survey indices are mean-standardized before use, and back-shifted to the start of the year (age-structured indices) or spawning time (biomass indices).
Estimated parameters	SURBA estimates all age effects of mortality, except for the oldest age (which is assumed equal to that for the next oldest age) and the user-specified reference age (which is assumed to equal 1.0). It estimates all year-effects of mortality, except the most recent which is assumed to be the arithmetic mean of the preceding three years). Finally, it estimates all cohort effects.
Objective function	The objective function is the sum of three weighted sum-of-squares: The sum-of-squared differences between observed and fitted age-structured survey indices; The sum-of-squared differences between observed and fitted biomass survey indices, if used; The sum-of-squared differences between subsequent estimates of the year-effect of total mortality. The latter SSQ is a penalty term intended to limit interannual fluctuations in mortality which may be driven by survey noise. The weight assigned to this term is provided by the user and is essentially arbitrary.
Minimisation	Least-squares regression minimization of sum-of-squares. Version 3.0 uses a NAG library routine to do this, but more recent implementations in ADMB and SAS (Noel Cadigan, DFO) use alternative functions.
Variance estimates and uncertainty	In Version 3.0, uncertainty estimates are generated for total mortality and recruitment only using the delta method. The R version now in development (SURBAR) uses sampling from a multivariate normal distribution derived from the inverse hessian to generate uncertainty estimates for all output quantities.

Other issues	Three more recent versions of SURBA are in development to address limitations of version 3.0: SURBAR (Coby Needle), SAS-SURBA and SURBA+ (both Noel Cadigan).
Quality control	SURBA was developed over a number of years in the context of ICES assessment working groups, and underwent extensive testing in these. It has also been considered at length by successive meetings of WGMG.
Restrictions	SURBA does not perform well with noisy survey data. In such cases, the uncertainty estimates become extremely wide, although it is not clear whether this is due to the method itself or the subsequent uncertainty estimates. Version 3.0 requires a full assessment dataset in the Lowestoft VPA format, although the more recent versions do not insist on this.
Program language	Fortran-90 with NAG and Winteracter libraries (SURBA 3.0), R (SURBAR), ADMB (SURBA+), SAS (SAS-SURBA).
Availability	SURBA 3.0 is held at ICES, and on many laptops around the world (it seems). SURBAR, SURBA+ and SAS-SURBA have not yet been released.

References	<p>Cook, R. M. (1997). Stock trends in six North Sea stocks as revealed by an analysis of research vessel surveys, <i>ICES Journal of Marine Science</i> 54: 924–933.</p> <p>Needle, C. L. (2002a). Preliminary analyses of survey indices for whiting in IV and VIId. Working Document WD2 to the ICES Working Group on the Assessment of Demersal Stocks in the North Sea and Skagerrak, Copenhagen, June 2002.</p> <p>Needle, C. L. (2002b). Survey-based assessments of whiting in VIa. Working Document WD1 to the ICES Working Group on the Assessment of Northern Shelf Demersal Stocks, Copenhagen, August–September 2002.</p> <p>Beare, D., Needle, C. L., Burns, F., Reid, D. and Simmonds, J. (2002). Making the most of research vessel data in stock assessments: examples from ICES Division VIa. <i>ICES CM</i> 2002/J:01.</p> <p>Needle, C. L. (2003). Survey-based assessments with SURBA. Working Document to the ICES Working Group on Methods of Fish Stock Assessment, Copenhagen, 29 January – 5 February 2003.</p> <p>Cook, R. M. (2004). Estimation of the age-specific rate of natural mortality for Shetland sandeels, <i>ICES Journal of Marine Science</i> 61: 159–164.</p> <p>Needle, C. L. (2004a). Absolute abundance estimates and other developments in SURBA. Working Document to the ICES Working Group on Methods of Fish Stock Assessment, IPIMAR, Lisbon 10–18 Feb 2004.</p> <p>Needle, C. L. (2004b). Data simulation and testing of XSA, SURBA and TSA. Working Paper to the ICES Working Group on the Assessment of Demersal Stocks in the North Sea and Skagerrak, Bergen, September 2004.</p> <p>Beare, D. J., Needle, C. L., Burns, F. and Reid, D. G. (2005). Using survey data independently from commercial data in stock assessment: An example using haddock in ICES Division VIa, <i>ICES Journal of Marine Science</i> 62: 996–1005.</p> <p>Needle, C. L. (2005). SURBA 3.0. Working Paper to the EU-FISBOAT WP3 Workshop, Rhodes, Greece, 7–11 Nov 2005.</p> <p>Cotter, J., Fryer, R., Mesnil, B., Needle, C. L., Skagen, D., Spedicato, M.-T. and Trenkel, V. (2007). A review of fishery-independent assessment models, and initial evaluation based on simulated data. <i>ICES CM</i> 2007/O:04.</p> <p>Needle, C. L. and Hillary, R. (2007). Estimating uncertainty in non-linear models: Applications to survey-based assessments. <i>ICES CM</i> 2007/O:36.</p> <p>Needle, C. L. (2008). Survey-based fish stock assessment with SURBA. Course given at the North-West Atlantic Fisheries Centre (DFO), St John's, Newfoundland, Canada, 3–4 September 2008.</p> <p>Mesnil, B., Cotter, A. J. R., Fryer, R. J., Needle, C. L. and Trenkel, V. M. (2009). A review of fishery-independent assessment models, and initial evaluation based on simulated data, <i>Aquatic Living Resources</i> 22: 207–216.</p>
Applications	<p>Used for many demersal stocks in ICES working groups for exploratory analyses, and for two stocks (VIa whiting and Norwegian coastal cod) to provide final advice.</p>

Annex 2.16: TINSS

Model & Version	Name and version number TINSS version 1.0 (there is no real version tracking yet).
Category	An age-structured model.
Model Type	TINSS is an age-structured model that is parameterized from a management-oriented approach. The leading parameters are MSY and Fmsy, from which the population parameters Bo and steepness are derived given age schedule information on selectivity, growth, maturity and natural mortality. The model is fit to data on relative abundance, age-composition and jointly estimates variance components for process errors and observation errors. Age-composition data are treated as a multivariate logistic observation and are weighted in the objective function using the conditional maximum likelihood estimate of the variance.
Data used	Minimum data requirements: Catch data (conditioned on historical catch information) Relative abundance Additional data that can be accommodated: Multiple abundance indices Age-composition data Length-composition data Mean age of the catch Mean weight of the catch Mean weight-at-age Environmental recruitment covariates Relative weights for survey data
Model assumptions	Assumes observation errors in relative abundance are lognormal Assumes logistic selectivity for both surveys and fleets, or age-specific selectivity coefficients for fishing. New cubic and bicubic spline selectivity options Fishing and natural mortality occur simultaneously, solving the Baranov catch equation using Newton's method Plus group aggregates individuals ages A and older and assumes constant mean weight-at-age for ages A and older. Maturity-at-age is assumed to follow a logistic curve and fecundity-at-age is assumed to be proportional to body weight. Growth follows a von-Bertalanffy curve, or empirical weight-at-age data are specified. Natural mortality is age-independent and time invariant. Informative priors can be specified for all model parameters.
Estimated parameters	Leading parameters consist of FMSY and MSY Selectivity parameters Natural mortality Annual recruitment deviations Growth parameter if length data available

Objective function	<p>cpue data (lognormal)</p> <p>Age-composition data (multivariate logistic)</p> <p>P(MSY) lognormal</p> <p>P(Fmsy) lognormal</p> <p>P(M) lognormal</p> <p>P(variance ratio) beta</p> <p>P(total variance) inverse gamma</p> <p>P(recruitment deviations) normal</p> <p>P(age 50% vulnerability) normal</p> <p>P(std in age at vulnerability) gamma</p>
Minimisation	Minimization is carried out using ADMB; its possible to estimate parameters in phases – this is controlled via a parameter control file.
Variance estimates and uncertainty	Variance estimates from either the inverse hessian, or the posterior samples constructed using the built in MCMC algorithm.
Other issues	
Quality control	
Restrictions	Currently restricted to a single fleet because the model is parameterized via MSY and FMSY that corresponds to that fleet.
Program language	ADMB
Availability	The source code and executable are available from the author <s.martell@fisheries.ubc.ca
References	<p>Forrest, R. E., Martell, S. J. D., Melnychuk, M. C., and Walters, C. J. (2008). An age- structured model with leading management parameters, incorporatig age-specific selectivity and maturity. <i>Can. J. Fish. Aquat. Sci.</i>, 65:286–296.</p> <p>Martell, S. J. D., Pine, W. E., and Walters, C. J. (2008). Parametrizing age-structured models from a fisheries management perspective. <i>Can. J. Fish. Aquat. Sci.</i>, 65:1586–1600.</p> <p>Martell, S. (2009). Assessment and management advice for pacific hake in u.s. and canadian waters in 2009. DFO Can. Sci. Advis. Sec. Res. Doc., 2009/021:iv+54p.</p> <p>Pine III, W., Martell, S., Jensen, O., Walters, C., and Kitchell, J. (2008). Catch-and- release and size limit regulations for blue, white, and striped marlin: the role of postrelease survival in effective policy design. <i>Canadian Journal of Fisheries and Aquatic Sciences</i>, 65(5):975–988.</p>
Applications	The model has been applied to Pacific Hake.

Annex 2.17: TSA

Model & Version	TSA
Category	Age-based
Model Type	<p>State-space model of a fishery, with the state vector giving numbers and fishing mortalities-at-age. The fishing mortalities-at-age are allowed to evolve through time.</p> <p>The model was originally formulated by Gudmundsson as a more rigorous statistical model of catch-at-age data than those then available. The TSA implementation was motivated by the need for a model in which the catch data could be excluded in particular years (e.g. when sullied by misreporting).</p>
Data used	<p>Catch-at-age (or discards and landings-at-age) with survey indices-at-age. Must have a full set of catches-at-age (or discards and landings-at-age) in the first year of the time-series, but subsequently missing years, or missing values within years are accommodated.</p> <p>Any number of surveys can be included, but cannot precede the catch data. Typically ages run from 1 to a plus group (in which case different stock–recruit functions can be used). Can be used with other age ranges, but options are restricted.</p>
Model assumptions	<p>Fishing mortalities-at-age are allowed to evolve over time both in terms of shape (i.e. selection) and overall level – both transitory and persistent changes are allowed.</p> <p>Survey catchabilities-at-age are assumed to have the same shape over time, but the level can evolve – both transitory and persistent changes are allowed.</p> <p>Fishing-mortality is assumed to be the same for all ages in the plus group.</p> <p>Maturity, natural mortality and weights-at-age must be specified, and can be age and year specific.</p>
Estimated parameters	<p>Numbers-at-age and fishing-mortalities-at-age (by year) with approximate standard errors.</p> <p>Also stock–recruit parameters, survey catchabilities, and variances that determine how much the fishing mortalities and survey catchabilities evolve over time.</p>
Objective function	<p>Negative log-likelihood is minimized.</p> <p>Catches-at-age are assumed to be normally distributed with constant cv.</p> <p>Survey indices-at-age are also assumed to be normally distributed with a (survey-specific) constant cv.</p> <p>Can downweight particular indices, or years, or ages that are particularly variable, or outlying.</p>
Minimisation	NAG minimization routine.
Variance estimates and uncertainty	Variance parameters are estimated by maximum likelihood, and the variance of the state vector (numbers-at-age / fishing mortalities at age) follows.
Other issues	Wide range of extensions possible (e.g. has separated fishing mortalities into components due to human consumption and industrial fleets), but time consuming and requires detailed knowledge of the underlying code.
Quality control	Extensive use over time, but no formal testing.

Restrictions	<p>Can't deal (easily) with sudden changes in fishing practice (e.g. a moratorium, or introduction of high grading)</p> <p>stock recruit relationships other than for age 1 recruits</p> <p>lots of zero survey indices (or catches)</p> <p>age zero catches or indices</p> <p>time-series that don't start with a full set of catch data</p>
Program language	Fortran (with R wrapper currently under development).
Availability	<p>Currently only available as poorly documented source code (and requires NAG libraries). Will soon be available as better-documented R function, but will still require NAG licence.</p> <p>From author.</p>
References	<p>Gudmundsson G, 1994. Time series analysis of catch-at-age observations. <i>Applied Statistics</i> 43: 117-126.</p> <p>Fryer RJ, 2002. TSA: is it the way? Appendix D in report of Working Group on Methods on Fish Stock Assessment. ICES CM 2002/D:01.</p> <p>and a whole heap of working documents to various ICES WGs.</p>
Applications	<p>Recently used for assessment of VIa cod and haddock and Faroese saithe – historically has been used to assess VIa whiting and IV cod, haddock and whiting.</p>