ICES WGMG REPORT 2011

SCICOM STEERING GROUP ON SUSTAINABLE USE OF ECOSYSTEMS

ICES CM 2011/SSGSUE:08

REF. SCICOM

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

10–19 October 2011

Vigo, Spain

ICES International Council for the Exploration of the Sea

CIEM Conseil International pour l'Exploration de la Mer

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

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

Recommended format for purposes of citation:

ICES. 2011. Report of the Working Group on Methods of Fish Stock Assessment (WGMG), 10–19 October 2011, Vigo, Spain. ICES CM 2011/SSGSUE:08. 250 pp.
https://doi.org/10.17895/ices.pub.9069

Contents

Executive summary

The ICES Working Group on Methods of Fish Stock Assessment (WGMG) met in Vigo, Spain during 10–19 October 2011.

ToR 1a: Data screening techniques prior to the selection of stock assessment models are useful for exploring and demonstrating data features, checking for consistency within and between data sources, providing ball-park trends to be expected from assessment model, and understanding behaviour of assessment models. It was felt that these techniques are not used enough, and it was recommended that showing outputs from data pre-screening techniques that proved informative should be a standard requirement in ICES stock assessment WG reports.

ToR 1b: Retrospective indices (based on e.g. Mohn's ϱ) were shown to have potential for developing threshold levels beyond which inaccuracy of assessment methods would be unacceptably large. Furthermore, the potential for using estimates of survey sampling variability as inputs to XSA to weight individual survey data points (by year and age) as a means to improve retrospective patterns was also explored. It is recommended that estimates of survey sampling variance always be calculated, and where appropriate, the inverse of survey estimates of sampling variance should be incorporated as a maximum weighting for corresponding survey data points.

ToR 1c: The move from simple to more complex assessment models is often motivated by the fuller use of available data/more biological realism (e.g. southern anglerfish), the fact that simpler models can give deceptively small confidence intervals, and the availability of more flexibility, for example to investigate the impact of changing the selection pattern (e.g. Celtic Sea and southern megrim), and more appropriate modelling of landings and discards (e.g. North Sea cod). However, moving to more complex assessment models has the danger of over-parameterization. Therefore, there is an overarching concern that "acceptable" model choice approaches are followed and model-fitting diagnostics are obtained (e.g. residuals are broadly random). Furthermore, although residual patterns may not be corrected for (e.g. autocorrelation), it is important to be aware of them, particularly in the context of MSE, to ensure that pseudo-data have the same properties as actual historic data. The Ecoknows perspective is that model specification should be driven by realistic biological and population dynamics assumptions and not data availability alone. It is recommended that consideration be given to using AIC in a frequentist or DIC in a Bayesian setting, for example, to guard against over-parameterization; and that when introducing random effects terms, the statistical properties assumed should be checked to the extent possible, e.g. when appropriate through a runs test to check for randomness.

A hind-cast/forecast simulation approach demonstrated (for North Sea plaice and sole) that harvest control rules which use fewer data (e.g. only survey indices of abundance) can outperform what actually happened in the past in terms of actual removals (based on a complex assessment) in almost every respect, particularly inter-annual variability in catch and fishing mortality. Furthermore, this MP testing framework could be used to evaluate the loss (in terms of more conservative catch limits) of reducing the amount of data collected. It is recommended that the approach used to evaluate simple management procedures, described in the report, be developed further as a possible framework for investigating the value of information.

ToR 1d: The investigations focused on year effects in surveys and estimating stock recruit relationships taking autocorrelation in recruitment into account by considering AR processes for residuals. The main conclusion for the two stocks investigated (3PS cod and American plaice) were that it was important to account for correlated errors to better reflect the information content of data; that better modelling of survey data should be carried out before asking assessment model to "figure it out"; that trying to estimate the 1st and 2nd order parameters of an auto-regressive process can lead to strange behaviour, requiring the imposition of a penalty to ensure residuals sum to zero; that there is a big difference in SR models estimated using the AR(1) or AR(2) formulations; and that there is potential that an autocorrelated recruitment error structure can confound the stock–recruit signal.

ToR 1e: No work presented explicitly addressing the topic of integrating uncertainty (although there are links to work elsewhere).

ToR 2: A review approaches for standardizing commercial cpue was provided, and an example GLM application based on the Tweedie distribution given.

ToR 3: Other ICES WGs have dealt with the topic of MSY reference points (e.g. WKFRAME, WKFRAME2, SGMAS). The study presented was limited to use of SURBA+ as the assessment model, and to 3PS cod and American plaice. Simulation analyses showed that measurement error in SSB, if substantial, could have a large impact on MSY reference points, and parameters such as S50% (SSB value at half asymptotic recruitment), and calculated from the estimated stock–recruit relationship. There is potential for providing guidelines for the use of more robust reference points, but further work is needed. With regard to the estimation of MSY when recruitment productivity varies, Bousquet *et al.* (2008) concluded that their study "reinforced the conviction shared by numerous researchers that biological reference points calculated in a deterministic framework can be far from optimal in stochastic settings". The study presented during the meeting found that the amount of process error had little effect on mean MSY reference points, which differs from the conclusions of Bousquet *et al.;* however upper and lower percentiles were affected. Furthermore, constraining multiplicative process errors to have a geometric mean of one makes a difference to results, but further work is need to draw firm conclusions.

ToR4: The WG commented on the categorization scheme proposed by SISAM, suggesting an additional subcategory reflecting possible management advice (e.g. types of reference points) under typical data for a given assessment category, and more detail on model assumptions. Concern was expressed that multispecies models were missing from the proposed 2013 symposium topics, but it was explained that multispecies models are not a main focus of SISAM, but that a session on such models would be included in the symposium. Suggestions were also put forward about datasets that could be considered for the symposium.

1 Introduction

1.1 Terms of Reference (ToRs)

The Working Group on Methods in Fish Stock Assessment (WGMG) chaired by José A. A. De Oliveira, met in Vigo, Spain from 10–19 October 2011 to:

1) In support of the ICES initiative on stock assessment methods, use simulation and case-study examples to help draw up guidelines for best practice when conducting stock assessments in the following areas:
    a) Data screening techniques prior to the selection of stock assessment models;
    b) Diagnostics to evaluate model fit (including measures of retrospective bias), and how these can be used to help refine models where appropriate;
    c) Guidance for deciding how complex a stock assessment model needs to be (e.g. how much to process/aggregate inputs; utility for advice);
    d) Implications and treatment of correlated errors;
    e) Integration of uncertainty (including accounting for retrospective patterns) in advice;
2) Review approaches for standardizing commercial cpue (available techniques and pitfalls).
3) Provide guidelines for calculating MSY reference points in a varying and stochastic environment.
4) **Comment on the proposed SISAM scheme for the categorization of assessment models.**

The ToRs were developed following during the WKADSAM meeting in September 2010 and subsequently agreed by SCICOM in October 2010.

1.2 Report Structure

A total of eight working documents (WDs) were presented to the meeting, and these are given, in full, in Annex 5. The report sections follow the ToR in order, with each section generally providing a summary of the presentations (referring to the WDs where appropriate), followed by a summary of plenary discussions. Where subsequent work was carried out, the plenary discussion summary is followed by the new material.

Southern horse mackerel data were made available to the group (see Section 1.3, and data folder of ICES SharePoint site), if a dataset was needed to help answer any of the ToR – it was subsequently only used under ToR 1a.

No work was presented under ToR 1e, apart from a presentation on the EU ECOKNOWS project, so the approach was to reflect discussions on how each of the presentations dealt with uncertainty, and how the ECOKNOWS project related to the various ToR.

The difficulty of estimating recruitment variability σ_R in an MLE setting, arose as a topic of interest, and Annex 6 describes the problem and presents some possible solutions.

1.3 Southern horse mackerel data

The Southern Horse Mackerel data were presented to WGMG as a possible dataset to be used, if one was needed, to help cover the various ToR for the workshop. Gersom Costas presented the data and assessment model applied, to help understand problems related to the data. Discards were not included in the assessment because they were considered low. The Spanish and Portuguese bottom-trawl surveys are combined and treated as a single index of abundance, with the Portuguese survey effectively receiving more weight because of the greater number of hauls. This is thought to be the only stock for which these two surveys are combined, on the basis of calibration studies that justified this – for other stocks, the two surveys are fitted separately, raising the question of why the surveys are treated differently for this stock.

2 Data Screening (ToR 1a)

2.1 Data screening techniques prior to the selection of stock assessment models

The presentation (Annex 5, WD 1) covered a review of existing plots used for data screening within and outside ICES. For catch data, bar charts and bubble plots are commonly used to display catch-at-age data, with bubble plots having the advantage of displaying cohort strength clearly. For cpue data, inter-age correlations (for the same cohort) provide information about the consistency of a survey with the same survey in other years. The index trends provide a guide to the biomass trends that can be expected from an assessment method. To examine mortality trends in an assessment-independent manner, catch curves can be plotted, and the negative gradient over a suitable age range gives the total mortality. Plotting spatial distributions of catch and survey data can indicate whether the survey design is appropriate.

Discussion:

When plotting cpue point estimates along with error bars, what exactly do the error bars mean (e.g. reflects sampling variation), and how is this accounted for in assessments when the survey only covers a portion of the stock distribution area? For the purpose of stock assessments, the estimates of uncertainty need to relate to stock abundance, and inclusion of additional variance over-and-above sampling variance, could help in this regard. It was also pointed out that the level of aggregation (by haul, by day, etc) is important when deriving estimates of uncertainty, because more aggregation can help with removing the problem of occasional large values, which can be difficult to handle from a statistical point of view (e.g. problem of distributions with heavy tails). Exploratory plots are meant as aids to decide on appropriate stock assessment model structure, and could help explain residual patterns in model diagnostic plots. It was suggested that consideration be given to mean-standardized plots (e.g. log-survey indices, where one subtracts the mean and divides by the standard deviation) over time to help see year and cohort effects.

Subsequent work:

Two case studies of preliminary data screening relating to southern horse mackerel and Celtic Sea cod are presented as an example of the kind of issues that can be raised at this point in the assessment process. Key plots from the working document are applied to these two stocks to illustrate the following aspects of the data:

- Age structure in the catch data
- Age structure in the index data
- Internal consistency of the index data
- Consistency between indices where more than one is available

Neither stock provided the opportunity to examine spatial trends in index or catch with the data available at this working group.

Southern Horse Mackerel

Data on southern horse mackerel includes catch-at-age from 1992–2010, for ages 0–11+, and a survey for the same years covering ages 1–11+. Figure 2.1.1 and Figure 2.1.2 show the catch data by age. In Figure 2.1.1, the bubbles indicate proportion of the catch-at-age in each year. In Figure 2.1.2, the catch is normalized on a log scale so that each age has mean zero, and standard deviation 1. This normalization is performed so that all ages are clearly visible, and the high variability typically present in recruits does not dominate the data. Both plots highlight the strength of the 1996 cohort, and the red bubbles in Figure 2.1.2 indicate that catches have generally been lower in recent years than the long-term average except at the lowest and oldest ages.

Figure 2.1.3 and Figure 2.1.4 show different ways of plotting the age structure of the survey data. Figure 2.1.3 is analogous to Figure 2.1.2, except that cpue is plotted rather than catch. The 1996 cohort that was clearly visible in the catch data are not prominent in the survey data, but there are clear year effects such as the high catch in 2005 across all ages. The strength of year effects is shown in Figure 2.1.4 by the degree to which the lines in the top left plot align. In contrast, if the top right plot shows a high degree of alignment, the survey would have strong cohort effects. The catch curves and gradient of the curves, shown in the bottom two plots, may indicate a slight decrease in total mortality over time.

Figure 2.1.5 shows the ability of the survey to track cohorts. In this case there are very few significant increasing correlations, so it can be concluded that the survey does not have a strong signal to noise ratio.

In this case study, the preliminary data analysis highlights the high level of year effects in the survey, and so any assessment will have to take account of the inconsistent catchability of the survey from year to year.

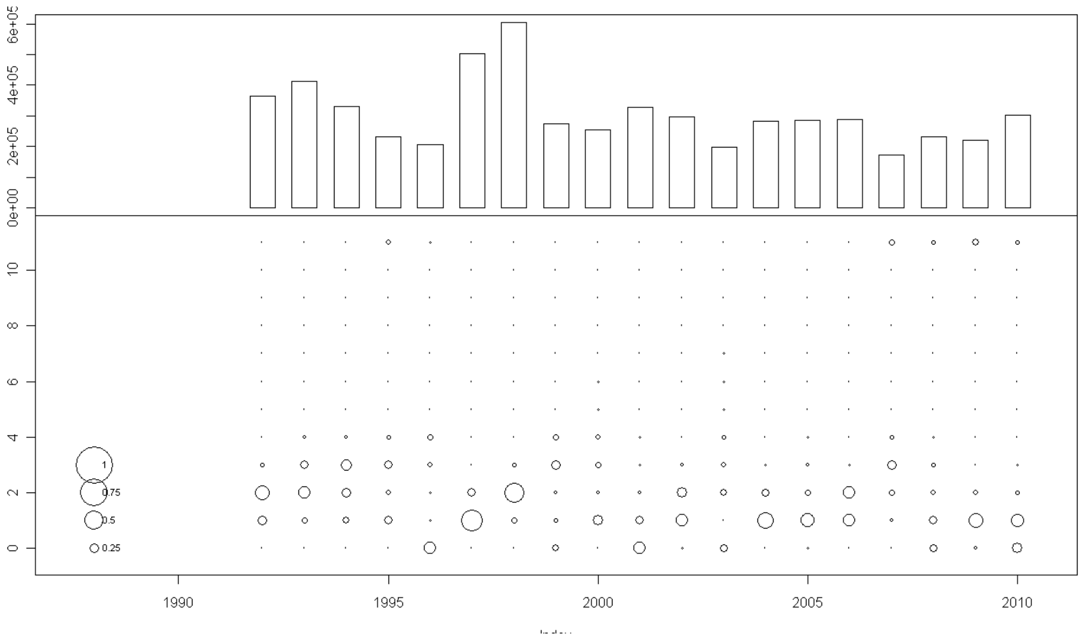


Figure 2.1.1. Landings by weight (top) and proportions of the catch numbers at each age (bottom) in the southern horse mackerel catch-at-age data.

Figure 2.1.2. Log catch-at-age data normalized so that each age has mean 0 and standard deviation 1. Black circles indicate positive values (log catch above average), red circles indicate negative values.

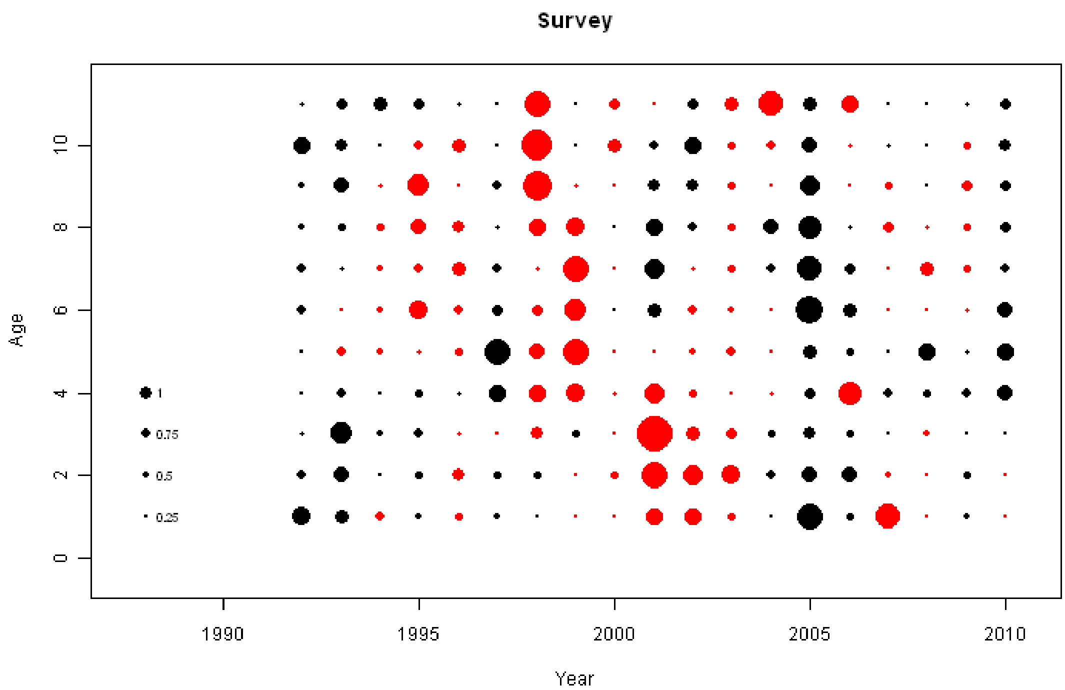


Figure 2.1.3. Log survey data normalized so that each age has mean 0 and standard deviation 1. Black circles indicate positive values, red circles indicate negative values.

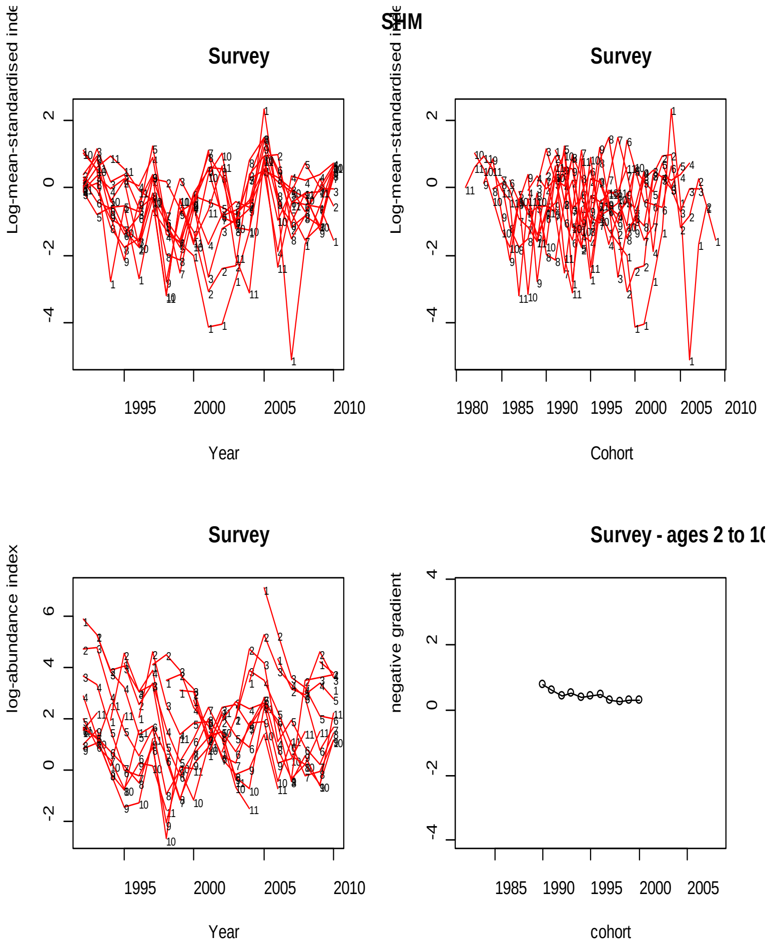


Figure 2.1.4. Southern horse mackerel survey showing index by age (top left), index by cohort (top right), catch curves (bottom left) and their average slope (bottom right).

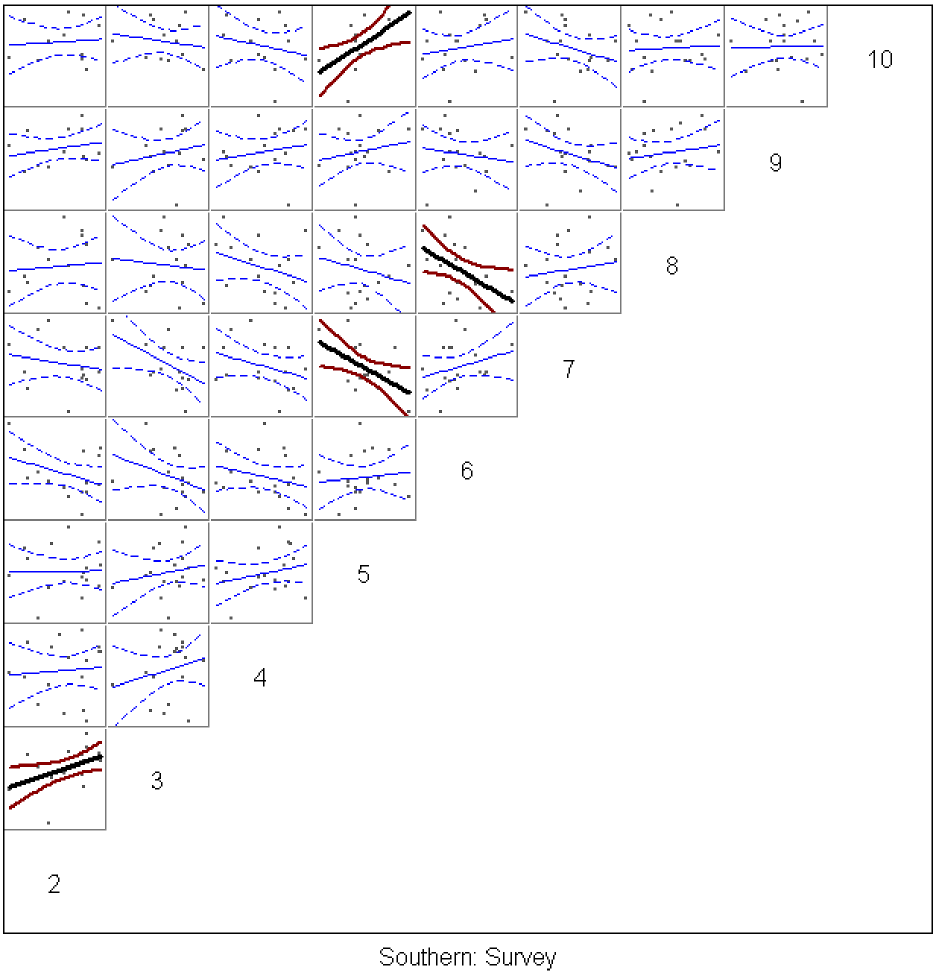


Figure 2.1.5. Internal consistency of the southern horse mackerel survey. Bold plots indicate significant relationships.

Celtic Sea Cod

Celtic Sea cod landing at age data are available for ages 1–10+ during the period 1971–2010. Eleven fleets area available for tuning, comprised of six series of commercial LPUE data, and 5 survey indices. This analysis focuses on one LPUE series (UK Otter trawlers in VIIe) and one survey (French EVHOE Groundfish Oct–Nov survey) as these have the longest time periods in common.

The landings in Figure 2.1.6 show that the majority of the catch is at age 2, accounting for at least 50% of the landings in many years. There are some strong cohorts such as the 1986 cohort visible from the proportions at age.

Figure 2.1.7, Figure 2.1.8 and Figure 2.1.9 illustrate LPUE data from the UK otter trawl fleet. Figure 2.1.7 shows strong cohort effects, and this is corroborated by the internal consistency of the LPUE shown in Figure 2.1.8. Figure 2.1.9 shows a decrease in LPUE, with no other strong year effects, but strong cohort effects. The catch curves do not indicate an increasing or decreasing trend in mortality. At present, the commercial indices do not include estimates of discards, but it may be more informative to plot cpue (i.e. landings and discards) than LPUE if these data are available.

Figure 2.1.10 and Figure 2.1.11 show data from the French EVHOE groundfish survey. Figure 2.1.10 suggests that there are no obvious year effects, but that there are clear cohort effects. In contrast to the data from the LPUE, the survey catch curves indicate a decrease in total mortality over time. The internal consistency, shown in Figure 2.1.11, of the survey is low, with only one significant correlation between years.

A comparison of the trends from LPUE and survey data are shown in Figure 2.1.12, which shows a strong correlation between the Survey and commercial LPUE for ages 1 and 3. For other ages, the correlation is weaker.

The pre-screening indicates that the survey may be useful for younger ages, but may not have sufficient data at older ages. It may be worth considering whether the different trends in mortality shown by the survey and LPUE can be investigated by looking at additional sources of data, to see whether catchability is changing over time for either index.

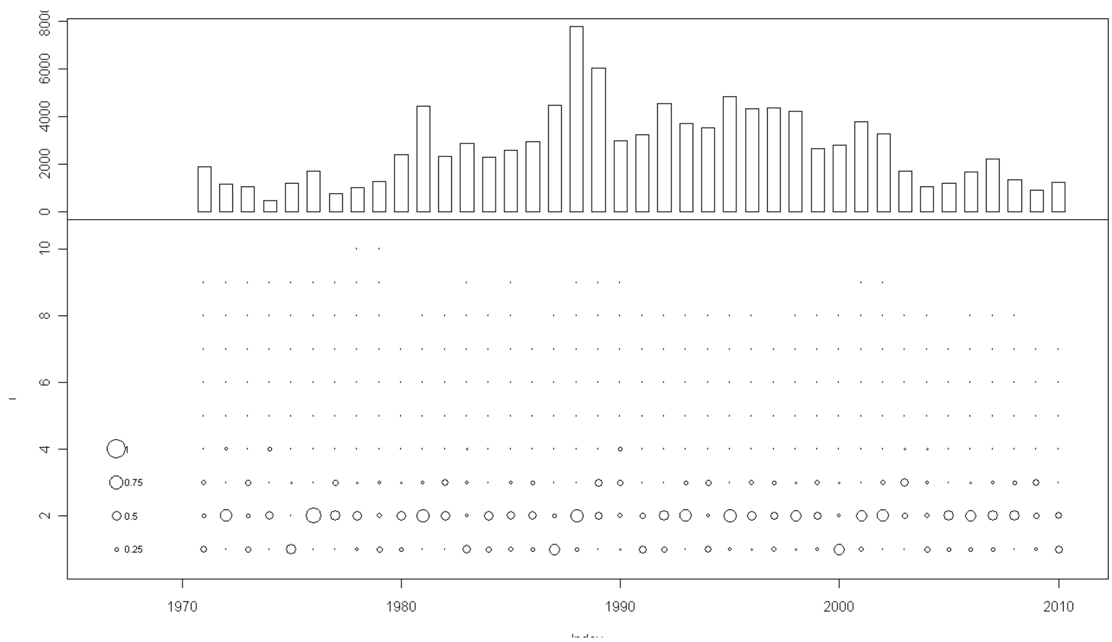


Figure 2.1.6. Landings by weight (top) and proportions of the catch numbers at each age (bottom) in the Celtic Sea cod catch-at-age data.

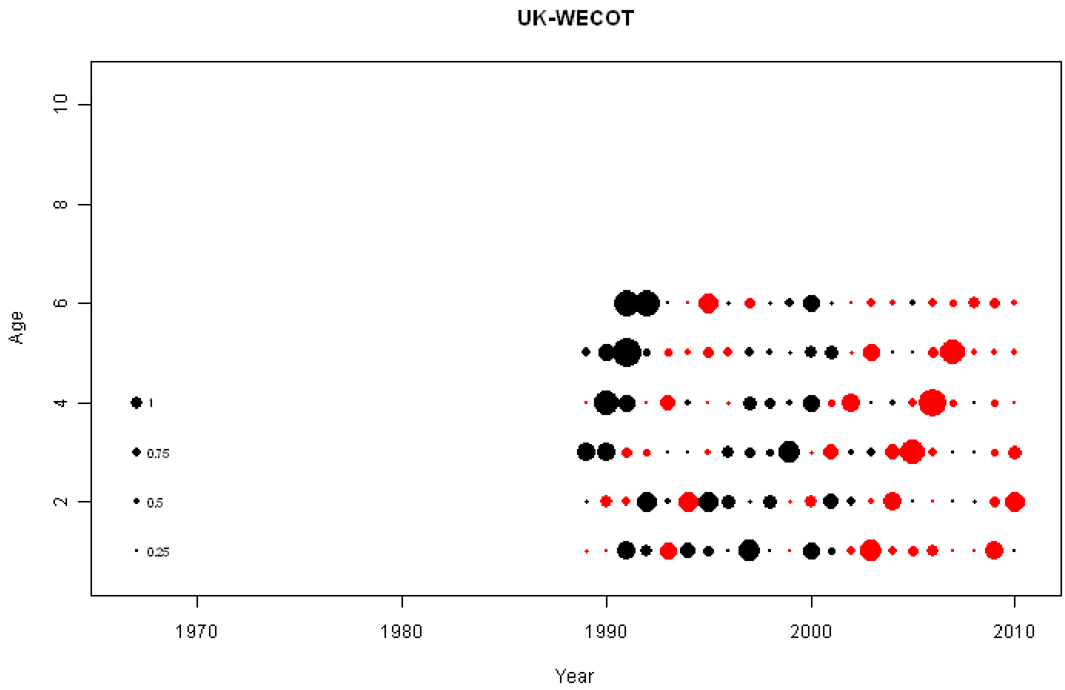


Figure 2.1.7. Log LPUE data normalized so that each age has mean 0 and standard deviation 1. Black circles indicate positive values, red circles indicate negative values

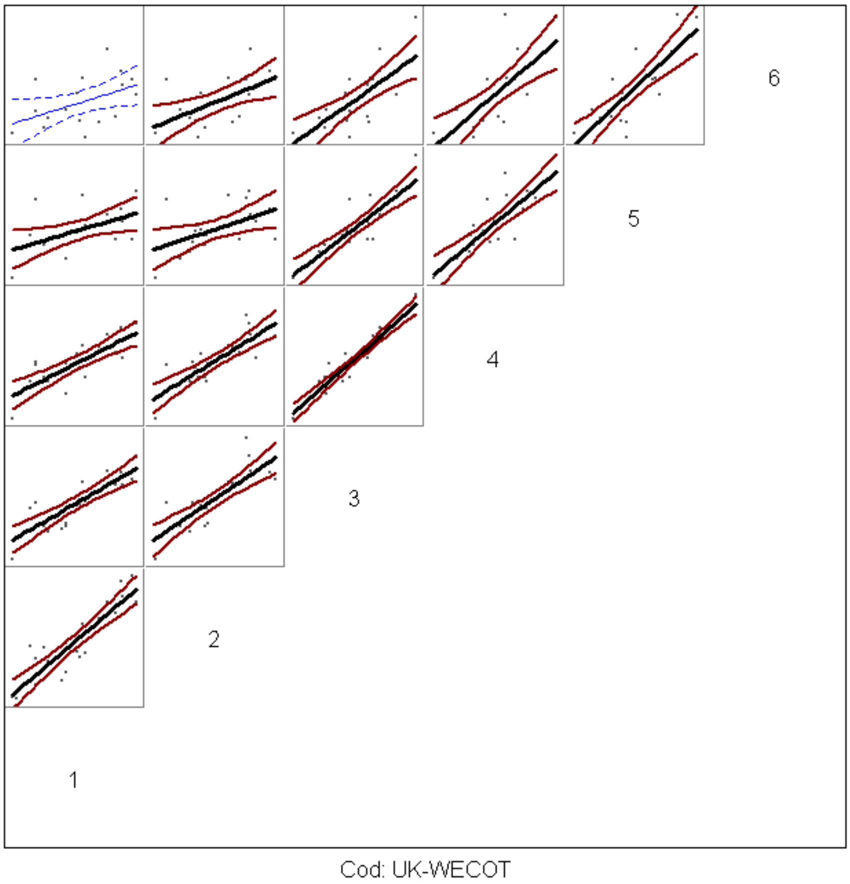


Figure 2.1.8. Internal consistency of LPUE data.

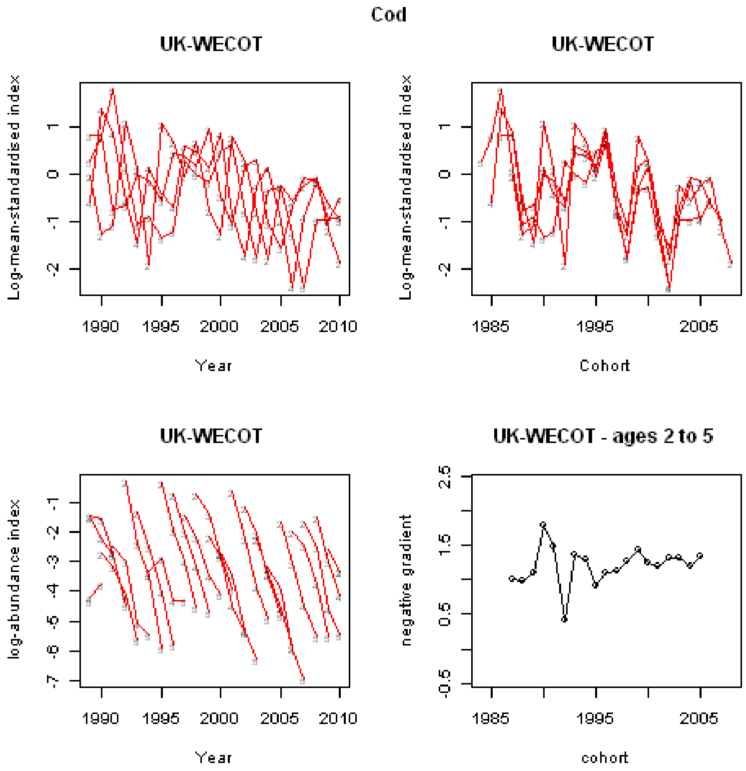


Figure 2.1.9. Celtic Sea cod LPUE showing index by age (top left), index by cohort (top right), catch curves (bottom left) and their average slope (bottom right).

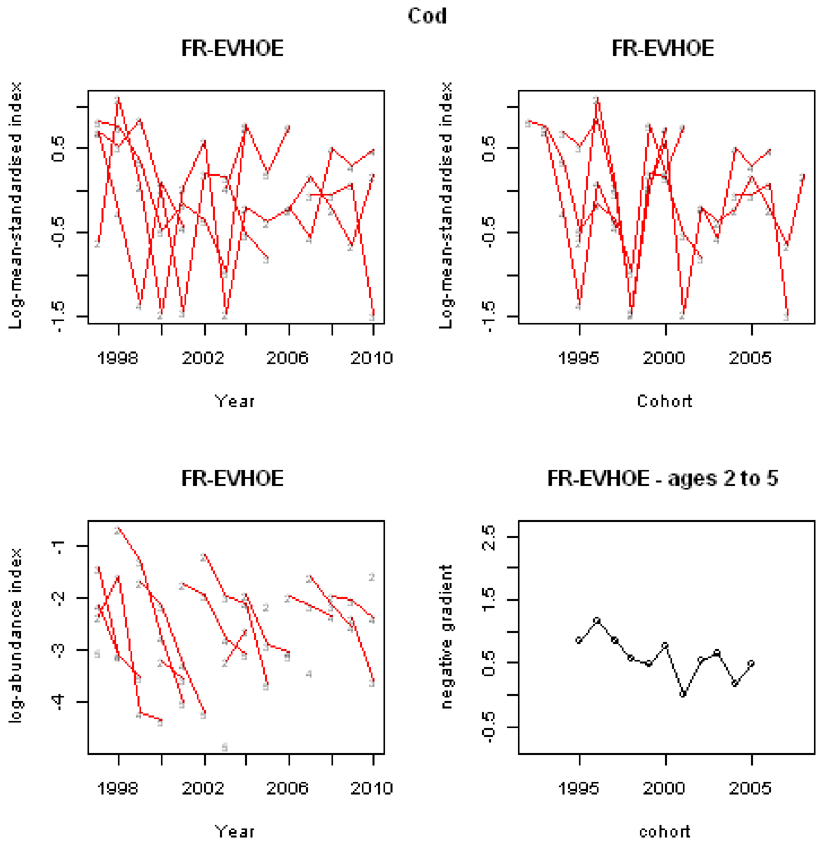


Figure 2.1.10. Celtic Sea cod survey showing index by age (top left), index by cohort (top right), catch curves (bottom left) and their average slope (bottom right).

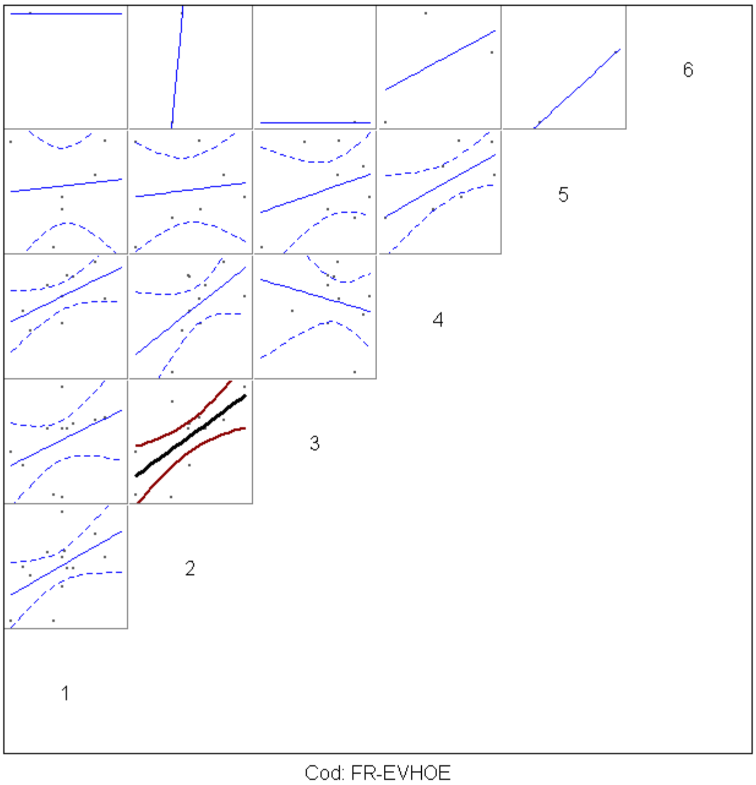


Figure 2.1.11. Internal consistency of the survey index.

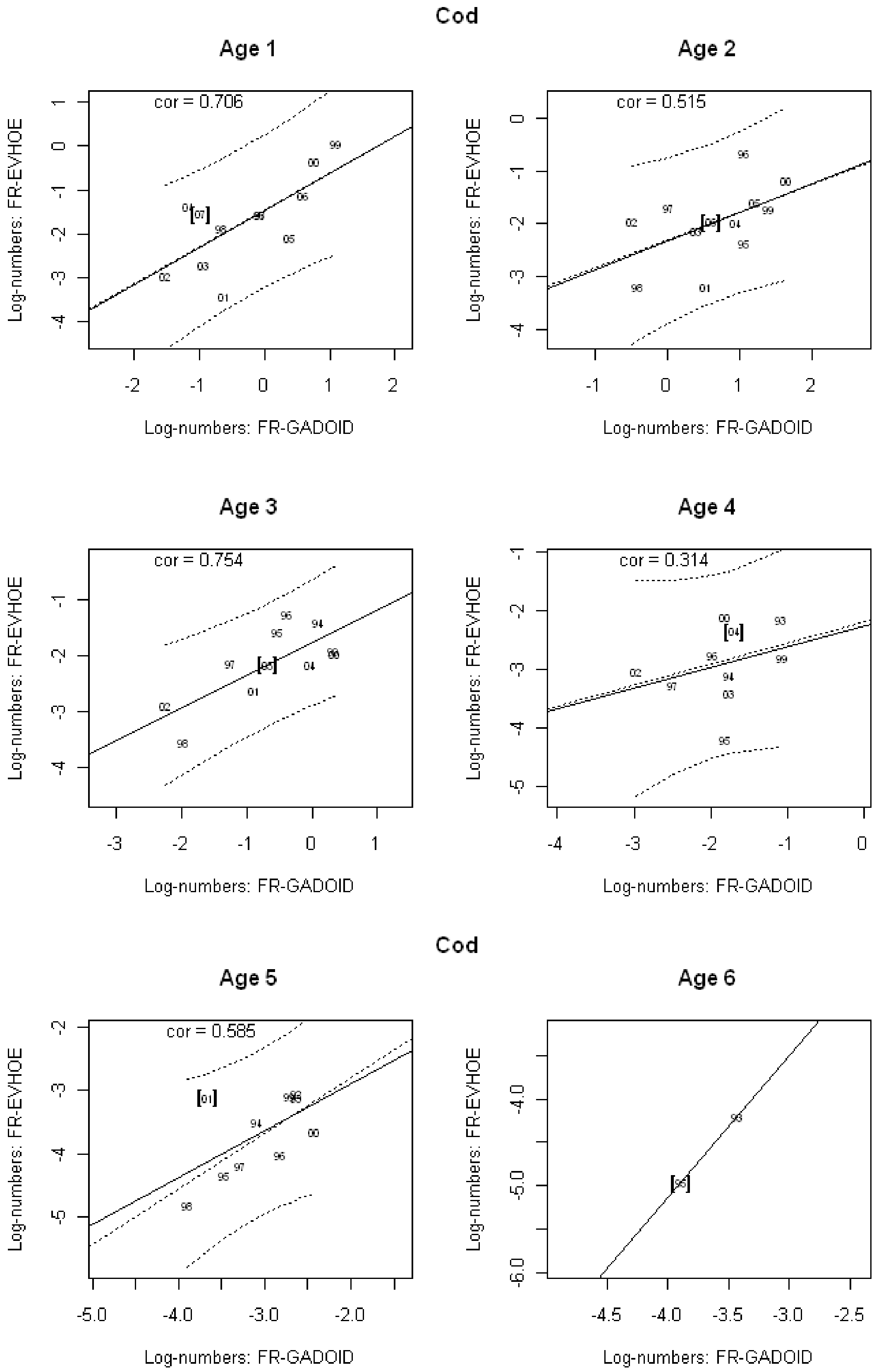


Figure 2.1.12. Consistency between the survey and LPUE data.

2.2 Assessment of Celtic Sea cod

The assessment of the Celtic Sea cod (VIIe-k), currently includes both commercial cpues and annual surveys indices. Surveys indices have always been based on a small number of cods (a few dozen to over a hundred individuals) because surveys in contrast to commercial fleets are not targeting cod and because they take place at the end of the year where fish are present on grounds that are not easily trawled. The use of those survey datasets has often been questioned at assessment or review groups because of the small number of fish.

However, exploratory assessments using SURBA show some relatively good cohort tracking and highlight the fact that survey data, as opposed to commercial cpues, are able to capture trends on recruitment; therefore their inclusion in assessments has been maintained over the years. The scaled weights of the different indices show that most of the information comes from surveys indices for young age classes and from commercial fleets for older age classes. Due to the design and period of the surveys, the probability of catching small cod is naturally higher than for commercial fleets which are catching bigger individuals with or without targeting that species.

Some exploratory assessments were carried out by combining or not surveys and commercial fleets. Estimates of recruitments are highly dependent on survey indices. Estimates of SSB are higher when both indices are used compared to using only commercial cpues. Trends on Fbar on ages 2 to 5 are more influenced by fleets than survey data, but fishing mortality is higher in recent years when using only survey data. Investigations suggest applying different weights and age ranges to indices whether they are from commercial fleets or from surveys.

Discussion:

There is the possibility of different selectivity between survey and commercial gears. Moreover the standardization of commercial indices and the use of commercial and survey indices for different ages in the model should be considered. The analysis of retrospective patterns was suggested. It was also suggested that pre-screening diagnostics (ToR 1a) would be useful as a means to check consistency within and between various data sources (see Section 2.1).

3 Diagnostics (ToR 1b)

3.1 Retrospective indices as a measure of bias in fish stock assessment.

Several indices were considered to measure disagreement among results of a retrospective analysis (Annex 5, WD 2). They were analysed by Monte Carlo simulation in response to random variability in partial recruitment, catch-at-age numbers and survey indices. The simulation also allows comparing VPA results with the original simulated values: several indices, called bias indices, are proposed to measure this inaccuracy in VPA results.

Properties of the retrospective and bias indices are analysed and, in particular, the relationship among them. Having noticed that some relationship exist between retrospective and bias indices, the potential use of the first index to infer a level of the second one was explored.

The relationship between each retrospective-bias index pair is not close: low retrospective indices do not imply low bias indices. Low retrospective indices are not a guarantee of goodness-of-fit. However some limits in retrospective indices could be established indicating unacceptable levels for some bias index.

As a case study, the best strategy for otoliths sampling was tested with the same Monte Carlo simulation. Catch-at-age and survey indices at age were calculated with the corresponding age–length keys made by simulated random sampling. It was concluded that a stratified sampling strategy is preferable to a random one. The effect of random error in age determination was also analysed. It was concluded that ageing error cannot be compensated by increasing sample size.

Discussion:

When pseudo-data were derived from the underlying simulation model, to be used when comparing fits to alternative assessment models, observed catches and survey indices were assumed to follow a lognormal distribution with fixed mean equal to the “true” values and varying CV in the range 0.0001 to 1.0. This lognormal assumption with fixed mean but varying CV implies that the median is not fixed, becoming smaller and further removed from the mean (and, hence, from the “true” value) as the CV increases. This inadvertently introduces a one side deviation in the observed data with respect to “true” values, which means that the metric used to measure retrospective bias is always positive. It was suggested that that either a normal distribution be used instead (but this potentially introduces the further problem of negative values for quantities that must necessarily be positive) or alternatively that the median instead of the mean be fixed at the “true” value (although the varying degrees of skewness of the lognormal distribution as its CV changes might still have an impact on the retrospective bias results). Another issue raised was that the lack of a retrospective bias does not guarantee that the model is any closer to the “truth”. Furthermore, the results are valid to the restricted case considered, and would need to span a much broader range of simulated population vs. assessment model combinations (including model misspecification options) to be able to draw meaningful conclusions. Other sources of retrospective bias could be tested for (e.g. due to changes in M).

3.2 Some reasons for retrospective bias in the stock assessment models

In a separate study (Annex 5, WD 3); consideration is given to one of the possible problems resulting in the retrospective bias seen in some stock assessments. The main idea is based on the use of statistical characteristics of abundance indices in the stock assessment model. The traditional approach assumes that for abundance indices, variance of the age groups is constant by years (XSA, ICA and others). But it is known (Heinmuth, Sparholt, Horbowy and others) this variance depends on the index value (dependence of log standard deviation as function of log index for age group). Results show that the relationship between abundance indices standard deviation and indices value is intrinsic to the abundance indices obtained on the basis of both bottom and acoustic surveys results, since they reflect the pattern of the Baltic fish spatial distribution (sprat, herring and cod).

This work shows, using as example the herring and sprat stocks in the Baltic Sea, the importance of the variance in the stock assessment. The Baltic international acoustic surveys data were used to calculate indices and their variances for these species. Quantifying and summarizing the main components of the overall uncertainty in sampling surveys was simulated using bootstrap re-sampling techniques and Monte Carlo simulation. We also developed a new version of XSA with weighted regression in the model and compared results with the traditional version. Application of the new XSA version considering the variability of abundance indices variance by years resulted not only in new estimates of fish stocks and population parameters (recruitment, total and spawning biomasses, mean fishing mortality rate), but also may change the temporal trends in fish stocks dynamics. The most important result from this work is that the values of retrospective bias are lesser in the modified approach as compared with the traditional approach.

Discussion:

A new version of XSA was presented that uses input variances that varied by year and age, which were derived from acoustic surveys using statistical and simulation methods. Questions were raised about whether an appropriate formulation of these input variances was used in the objective function of the new version of XSA. In particular, the CV of the untransformed survey indices, which typically is independent of the size of the survey index itself, would approximate the standard deviation of the log-transformed indices, so it is more common to use this CV (or a function of this: $\ln[CV^2+1]$) to represent the standard deviation of the log-indices in the objective function. This should be investigated further. It was suggested that the ability to weight survey observations individually already exists in some implementations of XSA, but this is rarely used in practice. It also exists in other stock assessment methods such as Stock Synthesis. The model has been tested with existing assessment data, rather than simulated data.

4 Model Complexity (ToR 1c)

4.1 Preliminary assessment of white anglerfish southern stock using Stock Synthesis (SS3)

A first attempt of assessment of white anglerfish southern stock using Stock Synthesis (SS3) is presented (Annex 5, WD 4) in order to evaluate its potential use as an alternative assessment model to the current surplus production model (ASPIC). Model structure, input data and provisional model settings are described in the work. Although more effort is required for tuning the model, the fit and the preliminary results seem to indicate that the Stock Synthesis can be an appropriate model to assess this stock.

Discussion:

Although the model needs annual total landings, any missing landings can be handled through specification of a prior with mean level based on expert opinion and large CV. The Beverton–Holt steepness parameter h is fixed at 0.999 and annual recruitment deviations estimated with high input CV (around 50–70%) for the corresponding priors. Since there is rarely enough information in the data to estimate h, and since the assumption of high h (close to 1) is a strong one, it was suggested that sensitivity to alternative values for h (say from 0.6 to 0.999), and its effect on model-fitting diagnostics, be explored. Selectivity functions have a 6-parameter double normal formulations, a form suggested as a starting point in the SS3 manual, but it was queried whether there was enough information in the data to estimate all six parameters, or whether a 3-paremeter logistic form (with the third parameter specifying a dome shape at the oldest ages) could do the job just as well. Verifying the type of selectivity may be difficult because the selection pattern is a combination of gear selectivity and the location of fishing activity.

It was pointed out that in SS3, it is easy to develop an over-parameterized model relative to the available data, because of the amount of flexibility allowed, so care is needed to strike the right balance between the amount of data added and the number of parameters estimated. Comparison between ASPIC (a production model) and SS3 (an integrated model using considerably more data) revealed different stock trends, but the SS3 model is still under development for this stock and needs fine-tuning. SS3 is due to be extended to allow weighting of length distributions by quarterly catches in order to allow the fitting of length data that are only available at the annual level - this would allow the SS3 model to be extended further back in time, making use of earlier data, thereby permitting greater contrast in trends. The move from ASPIC to SS3 for this stock was motivated by the unsatisfactory performance of the ASPIC model (e.g. biomass trends from ASPIC resemble trends in landings), and the desire to make better use of a greater range of data available for the stock (not possible under ASPIC; e.g. SS3 permits a retention ogive that allows discard data to be fitted and the incorporation of length structured data).

Subsequent work:

In an attempt to provide guidance for deciding how complex a stock assessment model needs to be, the process of migrating from a simple to more complex model was investigated. The idea was to analyse if a more complex model could provide a more realistic view of the population dynamic.

The particular example used to perform the comparison between a simple and a complex model was the white anglerfish southern stock (ICES Division 8c9a; Annex 5, WD 4). This stock is currently assessed by a non-equilibrium production model (ASPIC software) using as input data annual landings and two commercial LPUEs as abundance indices. Some concerns about the fit, especially due to its high dependence on landings information, seem to indicate that this model is not totally able to catch the reality of the population dynamic.

The decision to try a more complex model was also based on the possibility of incorporating more of the available biological information for this stock: length composition of landings, length composition of abundance indices (commercial LPUEs and survey), growth model and maturity ogive. Stock Synthesis v.3 (SS3; Methot, 2005) is an integrated assessment model whose main characteristics are its flexibility and its capacity to incorporate different types and sources of data. Also, it accounts for variation in fishing behaviour and the biological characteristics of an exploited stock over the history of the fishery. SS3 is a length-based integrated and statistical approach. Other advantages to apply integrated models instead of traditional assessment models are the possibility to introduce prior information.

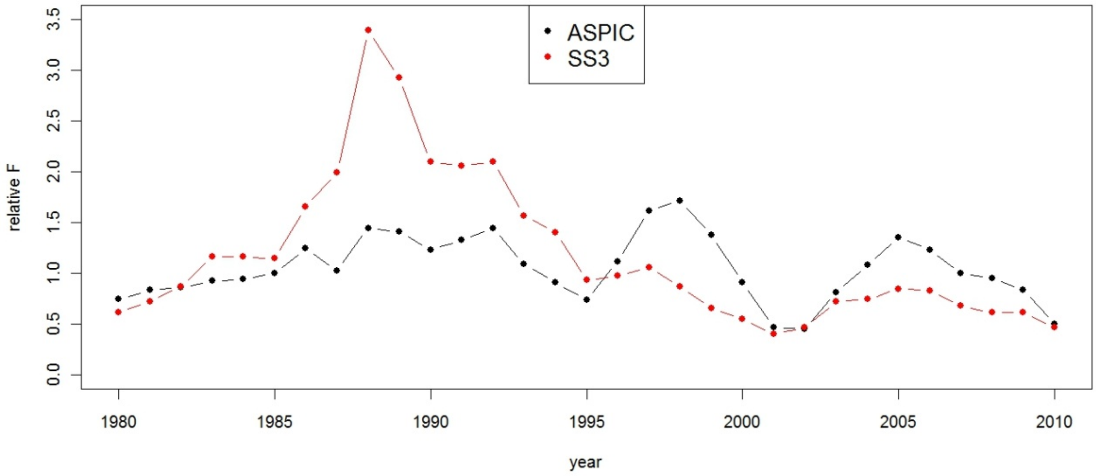
The SS3 model specifications were adopted according to previous stock information (based on literature and scientific work) and data availability. The adopted **time-step** for the SS3 configuration was quarter. Although an annual time-step could be employed, the quarterly time-step allows the model to represent potential seasonal processes such as recruitment.

The population dynamics was **spatially aggregated**: only one area was considered, but gear and geographical area have been used to define **fleets** with consistent characteristics (landings, size composition and selectivity). Based on these considerations, landings split into four fleets were defined, two fleets per country (Spain and Portugal).

The **size composition** data (for fleets and survey indices) were included in the model to potentially extract additional information about recruitment variability and fishing mortality. However, this implies additional assumptions about selectivity and representativeness of size sampling. Size-based models are an appropriate approach for stocks for which individual animals cannot be aged.

SS3 supports length-based **selectivity** with numerous functional forms. Different size-based double-normal selectivity curves by fleet were used. Each commercial LPUE, SPCORUTR8C and SPCEDGN8c, were introduce as four abundance indices (one by quarter) with a common selectivity curve. Erroneous assumptions about selectivity or poor selectivity estimates can result in incorrect estimates of stock abundance.

A Beverton–Holt stock **recruitment** relationship was assumed, with a fixed steepness of 0.99. In SS3 there is a single annual spawning biomass calculation, but the recruitment can be partitioned in various ways. For this case the recruitment was assumed to occur in the second (77%) and third quarter (22%), and recruitment in the second quarter is allowed to be time-varying.



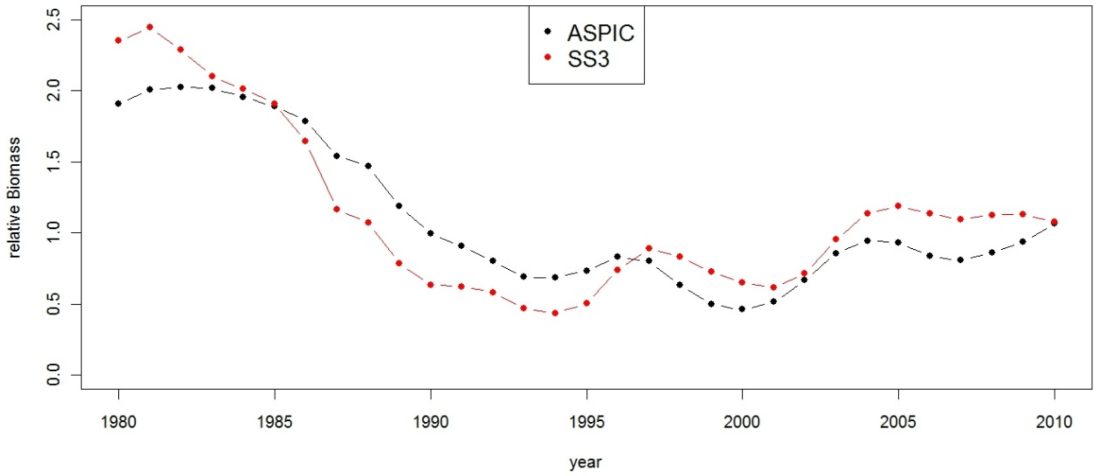


Figure 4.1.1. Comparison of relative estimates of fishing mortality (above) and total biomass (below) based on different assessment methods and data sources.

The cross-comparison of results from alternative models allows one to identify coincidences and patterns, being a guide for the detection of possible bias in data or erroneous assumptions. In Figure 4.1.1 a comparison of stock status outputs from ASPIC and SS3 are presented. Fishing mortality shows a similar pattern from both models at the beginning of the time-series; between 1985 and 1995, SS3 estimated higher values of F, with a maximum peak in 1988. The low values of F time-series were recorded in 2001 and 2002 by both models. Since 2005 F showed a decreasing trend in both model outcomes.

Total biomass for the non-equilibrium production and stock-synthesis models shows similar trends over the whole studied period. A decreasing period was observed at the beginning of the series (1980–1995) for both models. After reaching an absolute minimum values in 2000 (ASPIC) and a relative minimum value in 2001 (SS3), the biomass showed a very slight increase or stability in the last four years.

The analysis of the SS3 model fit and the residuals indicated that further investigation is needed to properly estimate the selectivity parameters of all fleets, and time variation in selection pattern for the Spanish artisanal fleet should be reviewed. However, the **over-parameterization** should be taken into account when using any complex stock assessment. The provided data may not be enough to estimate all the potentially important processes and the reduction of model bias can result in an increase of parameter estimation variance.

Although the SS3 model presented is still preliminary, the fit and results obtained indicate that SS3 provides an appropriate framework for integrating a diverse range of structural features and biological information for the assessment of the white anglerfish southern stock.

4.2 Applying a Bayesian model incorporating discards in the assessment of four-spot megrim (*Lepidorhombus boscii*) Southern stock

Since 2003 when the Data Collection Regulation started in the European Union, discards estimates have become more regularly available for an increasing number of stocks and fisheries. However, discards estimates tend to be patchy and are very often missing for earlier years. This has made their incorporation in stock assessments difficult. A Bayesian model incorporating the available discards estimates (while allowing for missing discards data for some years and/or fleets) was developed for the hake stock in ICES Divisions VIIIc and IXa by Fernández *et al.* (2010; Annex 5, WD 5). This age-based model estimates jointly the usual stock assessment quantities (population abundances, fishing mortality, etc) and missing discards. The presentation reviewed general aspects of this model and then showed a preliminary adaptation of it to the four-spot megrim stock in ICES Divisions VIIIc and IXa. This megrim stock is currently assessed by ICES with XSA (extended survivors analysis) without incorporating discards. However, discards are very significant, constituting around 60% of the total catch in numbers. Preliminary results from the Bayesian model were presented and compared with the XSA results from the ICES 2010 assessment. The main differences were that both recruitment and fishing mortality estimates for younger ages (the ages that are mostly discarded) were higher when using the model that incorporates discards. The megrim results were intended only as preliminary, and final tuning of the model for this specific stock is required before proposing it as a potential alternative to the current XSA assessment.

Discussion:

From a comparison between XSA and the Bayesian model, it was unclear why SSB trends were different, since the main difference is expected to be in recruitment and F for younger ages, given that the former model ignores discards, while the latter uses available discard information in an attempt to estimate missing discards. Furthermore, big differences were found for older ages belonging to the beginning of time-series. Apart from the use or not of discard information, these differences could be caused by differences in model formulation, such as differences in the treatment of the plus-group, and the fact that XSA uses tri-cubic weighting over 20 years (effectively down-weighting early data). It was suggested the Bayes model be run without discarding in order to compare with XSA results to understand which part of the difference in results was caused by difference in model configurations. In terms of whether or not to account for the effects of discarding, it was argued that it was important to consider discarding in the context of evaluating management plans in order to demonstrate the benefits of improving the selection pattern (e.g. to avoid discarding).

Subsequent work:

Summary of incorporating discards in assessments and why it is relevant in the content of management advice:

Discarding is a widely acknowledged problem in many fisheries. Although landings data have been collected for many decades, discard data in Europe generally have only been collected in the past 8 to 10 years. Discards estimates tend to be noisier than landings estimates, because they are based on fewer samples. The short discards time-series and their noise have hampered the incorporation of discards in stock assessments, which often implicitly assume that discards are zero. The impact that ignoring discards may have on assessment output and ensuing management advice is a source of concern and has been investigated by several authors (Casey, 1996; Williams, 2002; Punt *et al.*, 2006; Dickey-Collas *et al.*, 2007; Jardim *et al.*, 2010).

In a paper by Fernández *et al.* (2010) a Bayesian age-structured model was developed to take into account available information on discards and to handle gaps in the time-series of discard estimates. The model incorporates mortality attributable to discarding and appropriate assumptions about how this mortality may change over time are made. The result is a stock assessment that accounts for information on discards while, at the same time, producing a complete time-series of discard estimates. An earlier version of this model was presented at WGMG in 2008 (ICES, 2008a). The paper by Aarts and Poos (2009), a version of which was also presented at WGMG in 2008, deals with this same issue, albeit using different modelling assumptions and statistical fitting techniques. In the Fernández *et al.* (2010) paper, the Bayesian model was applied to the hake stock in ICES Divisions VIIIc and IXa, for which the available data indicate that some 60% of the individuals caught are discarded. Two runs of the model were performed; one assuming zero discards and another incorporating discards. When discards were incorporated, estimated recruitment and fishing mortality for young (discarded) ages increased, resulting in lower values of the biological reference points F_{max} and $F_{0.1}$, and generally, more optimistic future stock trajectories under *F*-reduction scenarios. It must be noted that this model is not currently used by ICES to assess the hake stock, due to the lack of an accepted age-reading criterion, so that a model that can use length-structured data (GADGET) is currently used by ICES for the assessment of this hake stock. In the current WGMG meeting, an adaptation of this model for a stock of megrim in ICES Divisions VIIIc and IXa was presented as a working document (Annex 5, WD 5) and during the meeting the model was applied to the megrim stock in ICES Divisions VIIb-k and VIIIabd, since age-structured data are available for these megrim stocks and both of them have substantial discards with many gaps in the discards time-series. The results obtained during the meeting are too preliminary for presentation in this report but the following general remarks (mostly extracted from the more detailed work in Fernández *et al.* (2010)) about the relevance of incorporating discards in assessments and their impact for management advice can be made.

The results obtained by Fernández *et al.* (2010) for the hake stock show that the main impacts of accounting for discards in the assessment are higher estimates of recruitment and *F* for the young ages (those that generate most discards), whereas the effect on estimates of SSB and $\bar{F}$ (the average of *F* over ages 2–5) is minor. The shift in exploitation pattern towards younger ages estimated when discards are accounted for leads to changes in the yield-per-recruit curve and lower estimates for the biological reference points F_{max} and $F_{0.1}$. This is in accord with the findings of other authors (Casey, 1996; Jardim *et al.*, 2010). The yield-per-recruit computations took into account the fact that yield consists only of the landed component of the catch. Chen *et al.* (2007) found that ignoring this (i.e. treating yield as the total catch even if part of the catch is discarded) may lead to overestimation of F_{max} and $F_{0.1}$. If a stock–recruitment relationship is fitted, including or ignoring discards in the assessment can also alter estimates of F_{MSY}. Williams (2002) found that ignoring discards led to overestimating the percentage of unfished SSB that would be retained under an MSY exploitation strategy (i.e. overestimation of B_{MSY}/B_0), a consequence of underestimating stock productivity in that case. No similar analysis was performed by Fernández *et al.* (2010), so no comparison can be offered.

A projection exercise under different hypotheses about recruitment and F was performed by Fernández *et al.* (2010) to explore the impact that including or ignoring discards in the assessment can have on short- and long-term stock prognosis. Three projection scenarios were considered regarding F:

1) F equal to the average over the final three assessment years in all projection years.

2) Starting from the values in the final assessment year, F decreasing by 10% every projection year, with the same reduction applied to all ages.

3) Starting from the values in the final assessment year, F decreasing every projection year by 30% for ages 0 and 1, and by 10% for older ages. This may reflect a situation where measures aimed specifically at reducing young fish mortality, e.g. closed seasons or areas, or gear modifications, are applied.

In all three scenarios, the probability that a fish is discarded when it is caught is age-dependent and was assumed to remain constant during the projection years, equal to that estimated by the assessment for recent years. Detailed results comparing the impact of including or not discards in the assessment can be found in Fernández *et al.* (2010), but their essence can be summarized as follows:

When no stock–recruitment relationship was considered (recruitment in projection years was drawn randomly from the posterior distributions of recruitment during a subset of the assessment years): Scenario 1 showed no differences in landings or SSB whether discards were incorporated or ignored in the assessment. Scenario 2 showed differences in long-term projections: incorporating discards in the assessment led to higher projected landings and SSB in the long-term. When discards were incorporated in the assessment, larger recruitment and F values were estimated for the young ages, so the benefits of reducing F became more apparent in the projections when discards were taken into account. Scenario 3 again showed that incorporating discards in the assessment led to higher projected landings and SSB in the long-term. In addition, when discards were incorporated, both landings and SSB were projected to be larger in the long-term under Scenario 3 than under Scenario 2, but no appreciable difference between Scenarios 2 and 3 could be noticed if discards were not incorporated in the assessment. Hence, comparison of Scenarios 2 and 3 also showed that when discards were incorporated in the assessment, it was possible to assess the effect of improving the fishery selection pattern (shifting it towards older fish), whereas this went undetected when discards were not considered in the assessment.

The consequences of changing recruitment assumptions are also of interest. Beverton–Holt and Ricker stock–recruitment models with lognormal recruitment deviations were also considered by Fernández *et al.* (2010) and fitted within the assessment model. Incorporating discards in the assessment again led to higher estimates of recruitment. Recruitment values in the projection years were then drawn from the stock–recruitment relationship including the lognormal departures. However, whereas for the range of estimated past values of SSB, the stock–recruitment relationships forecasted larger recruitment when discards were incorporated in the assessment, predicted recruitment from the stock–recruit relationship turned out to be smaller for the assessment that incorporated discards if SSB increased above that range of estimated past values. This had an impact on projections. As an illustration, long-term projections under Scenario 2 became more optimistic when discards were excluded from the assessment, in contrast to the result obtained under no stock–recruitment relationship. Fernández *et al.* (2010) concluded that great caution must be exercised when interpreting these results, given that the range of values of SSB and recruitment estimated for the assessment years does not permit inference of the magnitude of recruitment when SSB is considerably larger than what has been estimated for the past. The difficulty in estimating stock–recruitment models based on short time-series is widely acknowledged in the literature (see Brodziak and Legault, 2005). Hence, these projection results were regarded as speculative and shown only to illustrate that different conclusions can be reached depending on whether or not stock–recruitment relationships are assumed.

To summarize, coherently accounting for discarding in the assessment process is an important step for improving management advice. Having the discard mortality of the various fleets included as part of the model gives a wider range of scenarios that can be tried in projections. For example, the effect of reducing specifically the discard fishing mortality (achieved through, e.g. closed areas or seasons, or gear modifications), either for all fleets combined or for particular fleets, could be examined. This would permit a more detailed evaluation of the likely impact of a wider range of management options, provided that knowledge of the effect that a particular measure has on discard rates is available. As an example, the effect of gear modifications using results from scientific surveys designed to evaluate the difference in catchability between unmodified and modified gears could be explored.

Finally, as Aarts and Poos (2009) indicate in their discussion, if assumptions made about natural mortality are incorrect, part of the natural mortality may be incorrectly allocated to discards in the years with missing discard data, or vice versa. In such cases, discard estimates will most likely be incorrect, although it is possible that estimates of population abundance-at-age will still be correct if reliable relative indices of abundance-at-age exist. However, incorrect estimation of discard mortality may lead to incorrect conclusions being reached from the projections, for example when scenarios with an element of discard mortality reduction is considered.

4.3 Separating catches into landing and discards in the state-space assessment model used for North Sea Cod

The model currently used for North Sea Cod is summarized with special emphasis on the issue of estimating the so-called "unallocated mortality" in the last part of the data period (Annex 5, WD 6). The model is a state-space model, where the logarithm of the fishing mortalities are assumed to follow age-specific random walks. In the current model the unallocated mortalities enters the model via an estimated scaling applied to the catch in the relevant years, but in recent years it is believed that the discard estimates are the most likely origin for the mismatch between the signal from catches and survey information. The model is extended to use landing and discard as two separate data sources, such that the unallocated part can be assigned more correctly.

Discussion:

Diagnostics for model selection: Comparing the random effects model fits to fully parameterized statistical models:

- Number of free parameters in the model fit: There was the unresolved question about how many estimable parameters there were in such a model – this can be important in, for example, the use of AIC to compare different models. On the one hand, it could be argued that there was one only, namely sigma, as the others were integrated out. Alternatively, it could be argued that if sigma is prespecified, the count for a random effect as a parameter lies between 0 and 1, with those extremes corresponding to the limits of zero or infinite variance. However, the question arises in the situation where sigma is estimated within the model fit.
- Need to run checks for model-misspecification: It was agreed that runs tests on the outputs from the estimator for the random effects, if done correctly, would be a weak but reasonable test of whether they were indeed random. It was suggested that it is problematic to use maximum likelihood estimate of the variance when there are almost as many residuals as there are data points; perhaps need to use the bias-corrected REML estimate of variance instead. However REML is difficult to implement for non-linear models. A suggestion made was that the Laplace approximation frequently used to integrate out the random effects was essentially replacing the model by its localized linear approximation, and this might in turn allow for implementation of REML. This should be checked further - Justin Cooke and Hans Skaug might be able to advise.

Dealing with "unallocated mortality": black landings and discards

- It was suggested that it may be better to have separate equations for the landings and discards. This is because the scaling parameter ϕ which acts as the correction between landings and catch essentially sees catch playing a role as both a dependent and an independent variable (i.e. appearing implicitly on both sides of the regression equation).

Random effects models compared to Bayesian models:

- Difficult to distinguish between priors, likelihood and random effects.
- Are these models that different: It was suggested that Bayesian models to deal with discards in assessments (e.g. Annex 5, WD 5) are very similar to the random effects models presented here from a practical point of view (depending on the priors).

From a philosophical point of view, it was suggested that these were very different approaches: taken to the extreme, Bayesian analyses don't need any data to provide estimates as some statistic of posterior distributions, while random effects approaches use only data/observations and avoid any need for priors on fixed effect parameters. Therefore, the difference in approach depends on how "harmless" the priors are: one suggestion was that sensible priors are derived from the data, so why not rather include the data in the model fitting directly?

Stock–recruit relationship:

- It was noted that problems could occur when attempting to estimate the σ_R of the stock–recruit residuals because the maximum likelihood estimate often tends to zero with the fit to the S/R relationship dominating the likelihood. It was unclear whether the use of REML might resolve this problem. However it was reported that this had not been experienced in the random effects models implemented (Annex 5, WD 6), which estimated variance separately for recruitment. The problem is explored further in Annex 6.

Discussion following further updates (included in Annex 5, WD 6):

Two approaches were presented, the first a crude approximation that applies scaling to either the landings or discards instead of to the total catch, and the second a split model modelling landings and discards directly. Because of the difficulty of finding an appropriate way to handle zeros in the tight time constraint, the split model considers discard data at ages 1 and 2 only, and treats the relatively few zeros that still occur as missing data. The author was encouraged to further explore appropriate methods for handling zero data so that discard data for other ages could also be incorporated, since these discards have increased dramatically in recent years.

The difference in estimates of recruitment between the existing and split models, and the fact that the split model recruitment is often at the upper edge of the confidence limits for the existing model recruitment was noted. It was pointed out that recruitment is poorly determined and would be affected most by splitting the data. Recruitment is very sensitive to discards at the young ages.

The author was encouraged to provide good graphical diagnostic output, as this would help to understand what is going on, and to judge appropriateness of model fit. For example, it was speculated that discards have a high level of noise, and diagnostic output would help confirm this.

The fact that the crude approximation and split models are consistent in terms of SSB and average F(2–4) trends implies the former is a reasonable approximation. Nevertheless, the split model provides higher estimates of recruitment prior to the mid-1990s, which, given the consistency in SSB and average F(2–4) trends must imply higher Fs at the youngest ages for this approach.

The issue of the statistical appropriateness of the crude approximation, which came up in initial discussions above, was raised again. It was argued that applying the scaler to the discards in the simple way, as is done for the crude approximation, is much simpler than treating landings and discards as two separate data sources, and that the extra complexity doesn't seem to add much to the modelling approach. However, the crude approximation is conditioned on the landings fraction being "true", and given doubts about the reliability of landings and discards data, this assumption may not be appropriate. The author was encouraged to investigate alternative approximations (e.g. using a Taylor expansion) that dealt with this statistical issue.

The results presented applied the scaler over the whole period (i.e. since 1993), whereas applying the scaler to different periods and for different combinations (catch, landings only, discards only, etc.) could be investigated. However it was argued that any choice of how the scaler is applied should be done on a likelihood ratio test basis, and should rely on prior knowledge to justify the assumption that the source of bias had changed. It was suggested that one could also consider looking at separate multipliers on younger ages (dominated by discards) and older ages, but it was pointed out that this has been tried, and the capability already exists in the model, but that the approach was to develop the model in the direction indicated by this work.

WGMG supports the continued developments of the SAM model to deal with differences in scaling between landings and discards for North Sea cod.

4.4 An initial comparison of the performances of simple management procedures compared to complex assessments for some ICES stocks

These analyses (Annex 5, WD 7) aim to compare the fishery and resource consequences of management recommendations based on complex annual resource assessments to those based on simple empirical management procedures (MPs), which in the cases considered use only the annual abundance estimates from a single survey. The 2010 ICES assessments of the stocks of North Sea plaice and sole in Subarea IV are used for the investigation. The MPs are selected from the results of simulations based only on the resource information available in 1990. Their performances are then compared to what actually transpired over the 1990 to 2009 period under advice arising from the regular ICES assessments. For plaice, almost without exception the MPs' performances dominate what actually eventuated for every performance statistic: higher catches, greater final spawning biomass, lesser lowest spawning biomass during the 20 years, lower average fishing mortalities, and lesser interannual variation in both catch and fishing mortality. For sole these results are qualitatively duplicated, except for marginally smaller catches in some cases. In circumstances for ICES stocks where there may be difficulties in sustaining the level of sampling required for complex annual assessments, such as annual ageing of the catch, because of diminishing resources, these results are sufficiently promising to suggest that they be extended, in particular to further stocks, to confirm whether they might indeed provide a defensible alternative approach to the provision of scientific management advice.

Discussion:

The deterministic hindsight MPs for plaice are sensitive to the first two years, because XSA SSB and recruitment happened to decrease appreciably immediately after the start of projection period, causing projections to be sensitive to the high catch at that time. This results in the constant catch MP substantially outperforming what actually happened in terms of SSB, despite having only the initial two catches lower than the observed catches.

- The reasons for this behaviour need to be better explained in the document to help with understanding.
- As a check to test behaviour of the projections, it was confirmed that if the constant catch MP is replaced with realized catches in the hindsight deterministic projections, then the "XSA" estimates of SSB are matched exactly.

Robustness in biomass index assumption (i.e. a linear relationship with biomass, 0.2 SD lognormal error).

- Becomes relevant when checking if MPs based on such biomass indices can be implemented.
- For southern bluefin tuna assessments, based primarily on cpue (in contrast to survey) data, linear and power relationships between cpue and stock size are used; these are worth considering, but only if there is empirical evidence that is suggestive of them.
- What about systematic error? If plausible, there may be a need to include robustness trials that consider alternative relationships between the index and stock size, a breakdown in relationship (where the index has no relationship with stock size) or changes in the relationship over time. This would be the next step for this work, but where does one draw the line when it comes to potential sources of error to include (to a large degree some such suggestions could be pure speculation –

there needs to be some rule requiring some basis in data for suggestions made)? Standard practice has become to select from a larger set of plausible operating models (in this example, alternative hypotheses about the index vs. stock-size relationship) a subset of the most plausible ones, and to derive a weighted average of performance statistics across this subset, with the operating models within the subset typically being given equal weight.

Are projections realistic?

- For the two-line stock–recruit model the feeling was that projections were realistic (essentially based on the mean of past recruitment). It is merely the size of catches that have prevented this plaice stock from achieving its potential growth.
- Beverton and Holt with a steepness of 0.9 seems inappropriate for plaice.

How does performance relate to the quantity of data being used?

Here it is important to clarify that data are used both to condition the operating models used in testing MPs, and (when simulated into the future) to feed into the MP. This question relates primarily to the data input to the MP, which can either be used in an assessment model (here, XSA) that provides the parameters needed for the harvest control rule (HCR), termed "model-based" MPs, or it can be used directly in the HCR (in the study, the slope and target MPs), termed "empirical" MPs.

Previous papers (e.g. Punt 1993) have shown model-based MPs that attempt to take age-structure information into account (e.g. via VPA) when recommending catch limits led to greater variability in these limits than simpler age-aggregated production model MPs without any corresponding improvement in performance with respect to resource conservation. Further, for South African hake, the difference between model-based and empirical MPs has proved minimal in terms of resource conservation, and the simpler empirical MPs are more transparent, easier to check, and easier for stake-holders to understand, increasing the chance of buy-in. However, because model-based MPs were not explicitly considered in the study, the WG considered that the particular results of this simulation study are not sufficient to bring out this point clearly. Rather, what the study showed was that performance of the simple empirical MPs outperformed the combination of using a complex assessments AND meeting management objectives that changed over time.

Although MPs undergo a full review (including full testing) every 3–5 years and operate in "auto-pilot" in the interim, this does not mean that the pilot is missing, and "exceptional circumstances" provisions can be built in to dictate when to "switch off" the auto-pilot and call for an earlier review – this would happen when, for example, data fall outside the range over which the MP had been simulation tested. This approach would require the collection of enough data not just to service the MP itself on an ongoing basis, but also to be able to properly condition a reference set of operating models at least every 3–5 years, or when a full review of the MP is required.

The MP approach described above could potentially suffer if one were to stop gathering the data it requires. In order to evaluate the consequences of reduced data gathering (e.g. biannual ageing of catch-at-age data), both for conditioning operating models and for developing model-based MPs, the WG considered that a statistical catch-at-age model was needed, because it allows for missing data in a way that VPA-based models don't (e.g. missing catch-at-age data). Such an evaluation could already be performed using existing data by simply using a subset of these data for the MP testing, and comparing how the resulting increase in uncertainty would affect resource utilization for the same biological risk.

5 Correlated errors (ToR 1d)

5.1 Implication and treatment of correlated errors

This is a broad term of reference. Correlated errors in a general sense indicate systematic discrepancies between data and model predictions. Such discrepancies often indicate model mis-specification – transient or otherwise. The implications of correlated errors will depend on the magnitude of the correlations, where the correlation occurs, and the objectives of the modelling exercise. In particular, the implications of correlated errors will depend on the objectives of the stock assessment.

In the standard ICES assessment model (XSA) and in other traditional VPA-type assessment models (e.g. ADAPT), surveys and other tuning indices are assumed to be the primary source of errors. Errors in catch statistics and other biological data (e.g. weights, maturities, etc.) are assumed to be negligible compared to the errors in tuning indices. In this case, correlated errors will be evidenced by clusters of positive or negative survey residuals. The clusters will be defined in age and or time, and within or across tuning indices. Such correlations usually indicate that the model is over-simplified and that the resulting statistical inferences may be imprecise. Of particular concern is the potential for biased estimates of stock size and trends, and fishing mortality rates.

If the correlated errors are such that there is a trend in survey residuals over time then another implication of such errors is a retrospective pattern. Such patterns are usually associated with a trend in residuals.

In more modern stock assessment models that consider errors in other assessment inputs, primarily catches but possibly also weights at age etc, correlated errors may occur in these other data in addition to tuning indices. However, the basic implications are similar to the XSA case, and that is potentially biased estimates of stock size and trends, and fishing mortality rates.

Two common examples of correlated errors involve survey year effects and autocorrelation in residuals from a stock–recruit analysis. These are illustrated using a stock assessment model for cod and American plaice in NAFO Subdivision 3Ps. The assessment model is "developmental" and is described in Annex 5, WD 8. It is an extension of the SURBA model (called SURBA+) that provides estimates of total mortality rates and trends in stock size based only on tuning indices and other biological information. It is a catch-free assessment model. These correlated errors issues are generic, and not specific to the SURBA+ model.

Survey year effects

These refer to large increases or decreases in survey catch rates for all or most ages. This could also apply to cpue tuning indices. The magnitude of the changes is well beyond what is possible in the stock. Year effects are indicated by correlated survey residuals in which the residuals for a tuning index have the same sign for all or most ages in a year. Year effects are common in many stock surveys, and in particular for cod in NAFO Subdivision 3Ps.

The most recent assessment for this stock was based on SURBA+ (see Annex 5, WD 8). A modification currently explored is to split total mortality into a user supplied *M* component and an estimated fishing mortality (*F*) component. A separable *F* model is used. Survey residuals (Figure 5.1.1) indicate substantial year effects. Such year effects have also been noted in VPA-type assessments of this stock. These year effects create problems when modelling the between-year variation in the F year effects in SURBA+ and modelling recruitment residuals when a stock–recruit model is incorporated into the SURBA+ model.

Myers and Cadigan (1995) proposed a simple approach to accommodate year effects. The basic survey observation equation is modified as

$$I_{ay} = q_a N_{ay}^{(f)} \exp(\varepsilon_{ay} + \tau_y),$$

Where I_{ay} is the survey index for age a in year y, $N_{ay}^{(f)}$ is population numbers at the time of the survey in fraction of year (f), q_a are survey catchability coefficients, τ_y are independent $N(0,\sigma_Q^2)$ random year effects, and ε_{ay} are additional independent $N(0,\sigma_\varepsilon^2)$ observation error terms. We use this approach with the 3Ps cod SURBA+ model. Parameters were estimated via marginal maximum likelihood. AD Model Builder (ADMB Project 2009) is used to implement the model and estimation.

Random year effects resulted in a substantially better fit. The negative log-likelihood (nll) decreased by 54.7 which is large decrease for one additional parameter (σ_Q^2). Residuals look well-behaved (Figure 5.1.2) and the F random walk standard deviation decreased from 0.74 to 0.38 with a smoother trend in F over time. However, the estimate of σ_Q was 0.5, which seems too large. The predicted random year effects (τ_y) displayed some unexpected behaviour (Figure 5.1.3). The large year effects in 1990 and 1991 were unexpected because residuals for these years did not indicate strong year effects (Figure 5.1.1). The large year effects in 2009 and 2010 will be controversial because they indicate that the recent increases in survey catch rates are not proportional to the same increases in stock size. Really surprising was the trend in year effects throughout the 1980's and early 1990's. These year effects may be masking a real trend in the stock.

The large estimate for σ_Q suggests that the survey has limited utility for tracking the population. If there is no change in stock size between two years then year effects with $\sigma_Q = 0.5$ mean that the standard deviation of the log difference between two surveys is $2^{-1/2} = 0.71$, and that only large changes in catch rates can be interpreted with confidence to indicate the same changes in stock size. For example, the probability of getting greater than a 50% increase in catch rates due to year effects when stock size is constant is 0.28 (i.e. $N(0, 2^{-1/2})$ probability of exceeding log(1.5)). The probability of getting greater than a 100% increase in catch rates is 0.16. These are controversial results. The value of σ_Q was therefore fixed at 0.25, which is still a large value but indicates the survey is still useful for tracking trends in the stock. For example, if $\sigma_Q = 0.25$ then the probability of getting greater than a 100% increase in catch rates is only 0.025.

When $\sigma_Q = 0.25$ the reduction in nll (compared to the SURBA+ with no year effects) was 43.9 which is still large. The predicted τ_y's (Figure 5.1.4) were not as controversial compared to Figure 5.1.3 and residuals (Figure 5.1.5) did not indicate substantial year effects.

The treatment of year effects had some impact on estimates of SSB, particularly prior to 1990 (Figure 5.1.6). Estimates of F and recruitment (Figure 5.1.7) were also sensitive in some years to the treatment of year effects.

Summary

The 3Ps cod case study illustrates that year effects can be difficult to accommodate. Modelled random year effects may have trends for subsets of years, and there is a possibility that including year effects in a stock assessment model may mask real changes in the stock. We should treat trends in predicted year effects with scepticism, and require that a good explanation be provided for such trends before deciding that a trend in survey catch rates is not related to a trend in stock size.

There may be within-survey signals for year effects (e.g. temperatures, survey timing issues, coverage problems, other species, etc). These might shed some light on what are really year effects vs. true stock-signals. There is also potential that better modelling of the survey data could produce better survey indices with less year effects. It is recommended that these survey analysis issues be fully explored before asking an assessment model to sort it all out!

Autocorrelated stock-recruit residuals

This is a common phenomenon that indicates systematic discrepancies over time in stock recruitment productivity compared to that predicted by the stock–recruit model. Such discrepancies could be driven by persistent environmental or predator fluctuations, among other reasons.

American plaice in NAFO Subdivision 3Ps provides an illustration of this problem. A SURBA+ model is being developed for this stock, in part because age-composition information for commercial catches has not been available for recent years. A Beverton–Holt stock–recruit model can be fitted internally within the SURBA+ model,

$$R(S) = \frac{\alpha S e^{\varepsilon_{SME}}}{\beta + S e^{\varepsilon_{SME}}} e^{\varepsilon_{RME}}$$

where ε_{SME} is the measurement error (ME) in SSB and ε_{RPE} is the process error in the stock–recruit model. Note that the ME in SSB includes errors due to, for example, changes in sex ratios, changes in fecundity related to changes in the age-distribution of the SSB, etc. Both errors are assumed to be normally distributed with standard errors σ_{SME} and σ_{RPE}. These parameters are highly confounded (Annex 5, WD 8), so for illustration purposes σ_{SME} was fixed at 0.25 and σ_{RPE} was freely estimated. Note that although SSB ME can affect stock–recruit parameter estimates, it has negligible effect on stock–recruit residuals. AD Model Builder (ADMB Project 2009) provides empirical Bayes predictions of the ε's.

There is some indication of autocorrelation in the SURBA+ predictions of ε_{RPE}'s (Figure 5.1.8), with short periods of positive and negative residuals. The ar() R function was used to diagnose the autocorrelation structure in these residuals. The results indicated an AR(2) model with coefficients 0.56 and -0.32 providing the best description of these errors. Note that the same procedure was used then $\sigma_{SME} = 0$, and the results were nearly identical.

The recruitment residual likelihood component of the SURBA+ model was adjusted to be an AR(2) model, with the first two residuals assumed to be independent and identically distributed. The AR(2) coefficients and variance parameters were estimated via marginal maximum likelihood. The results were unexpected. The AR(2) recruitment residuals were poorly behaved and were not centered about zero. To fix this problem a penalty term was added to the nll fit function to penalize against this behaviour. The penalty function strongly encouraged the recruitment residuals to add to zero. The resulting recruitment residuals were much better behaved. Stock size estimates did not change much (Figures 5.1.9 and 5.1.10), but the estimates of the AR(2) coefficients (i.e. 0.90, -0.54) were considerably different from the external results above (i.e. 0.56, -0.32).

An AR(1) recruitment residual model was also explored. The AR(1) model resulted in only a slightly worse fit, but the stock recruit fit looked poor. The model was essentially flat, and the stock–recruit trend appeared in the residuals.

Summary

Accounting for autocorrelation in stock–recruit residuals internally within the SURBA+ model had little effect on assessment model output, with the potential exception of recent recruitment estimates.

Autocorrelation in stock–recruit residuals may be important to account for in short and medium term stock forecasts, and also in long-term forecasts and when deriving MSY reference points.

There is potential that an autocorrelated recruitment error structure can confound the stock–recruit signal. How to model the autocorrelation error structure in recruitment residuals requires further investigation

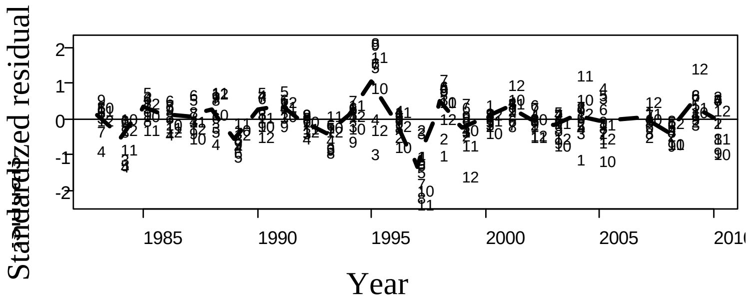


Figure 5.1.1. Standard residuals vs. year. 3Ps cod SURBA+ <u>without</u> year effects. The dashed line indicates the average residual each year. Plotting symbols indicate age.

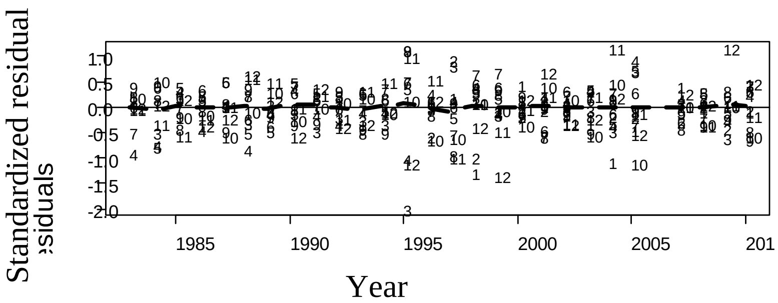


Figure 5.1.2. Standard residuals vs. year. 3Ps cod SURBA+ <u>with</u> year effects. The dashed line indicates the average residual each year. Plotting symbols indicate age.

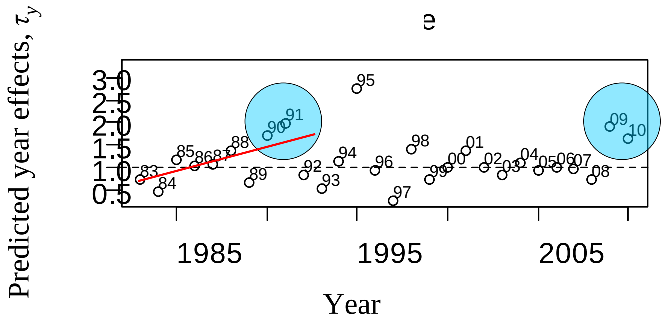


Figure 5.1.3. 3Ps cod SURBA+ year effects when σ_Q is <u>estimated</u>.

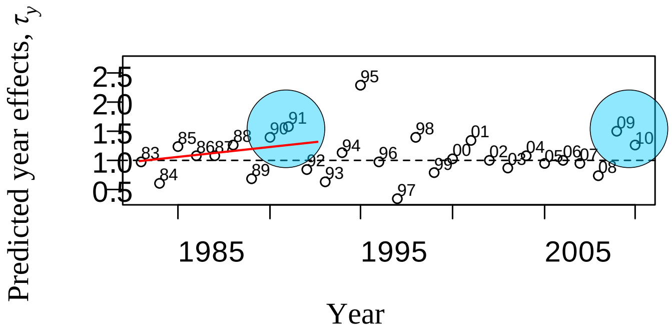


Figure 5.1.4. 3Ps cod SURBA+ year effects when σ_Q is <u>fixed at 0.25</u>.

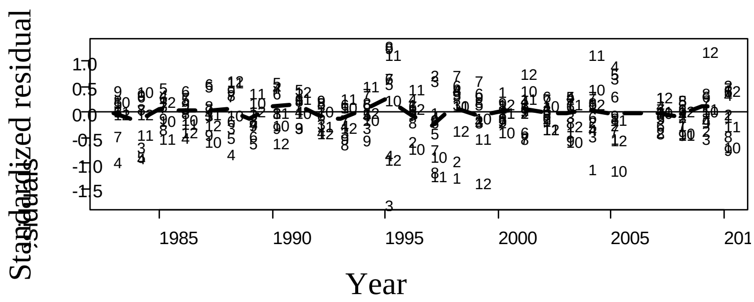


Figure 5.1.5. Standard residuals vs. year. 3Ps cod SURBA+ <u>with</u> year effects σ_Q fixed at 0.25. The dashed line indicates the average residual each year. Plotting symbols indicate age.

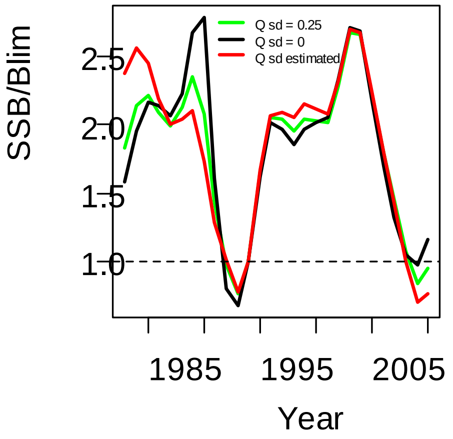


Figure 5.1.6. Trends in SSB relative to Blim (Brecovery – SSB in 1994) for 3Ps cod. Results are from the SURBA+ models with no year effects (black line), year effects variance estimated (red line), and year effects standard error equal to 0.25 (green line).

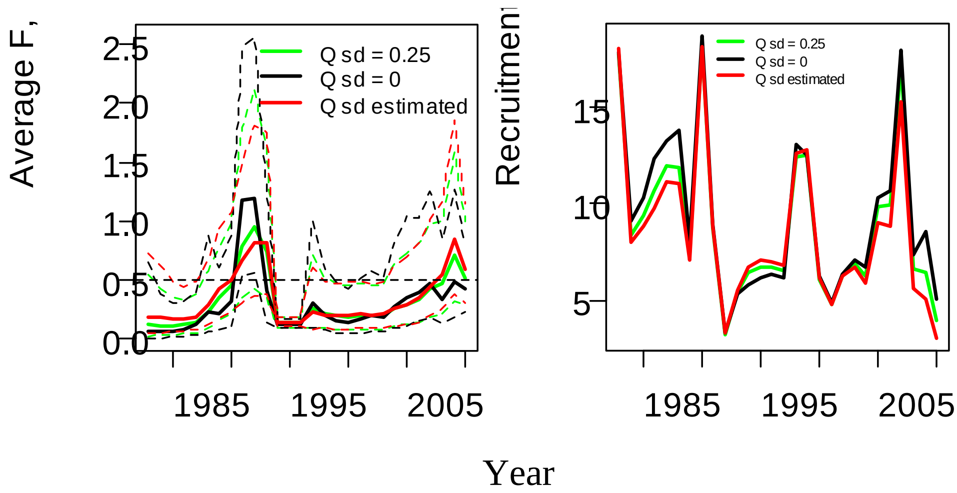


Figure 5.1.7. Average F with 95% confidence intervals (left) and recruitment (right) from three SURBA+ models. See Figure 5.1.6 for other details.

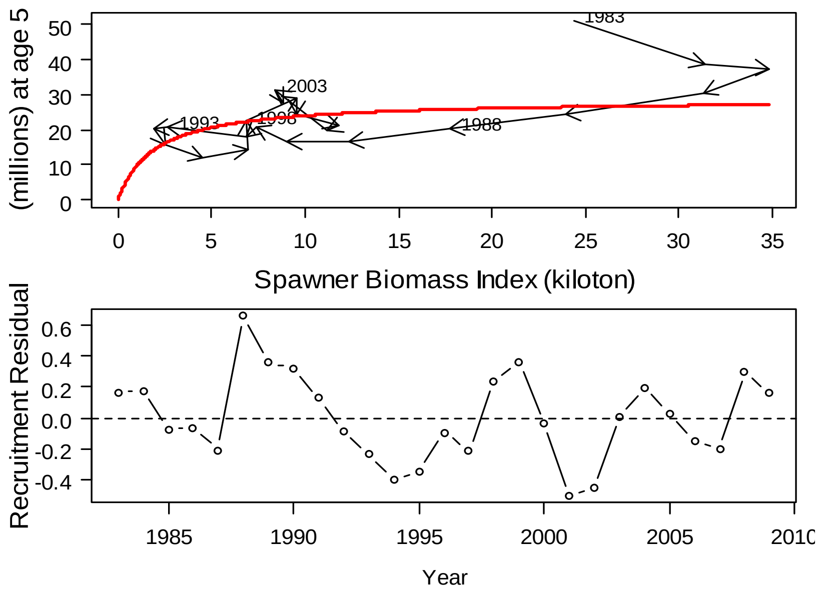


Figure 5.1.8. 3Ps American plaice stock–recruit results. Top panel: estimated Beverton–Holt model (red line) and predicted recruitment (arrows), connected by year with some years indicated. Bottom panel: time-series of recruitment process errors, ε_{RPE}.

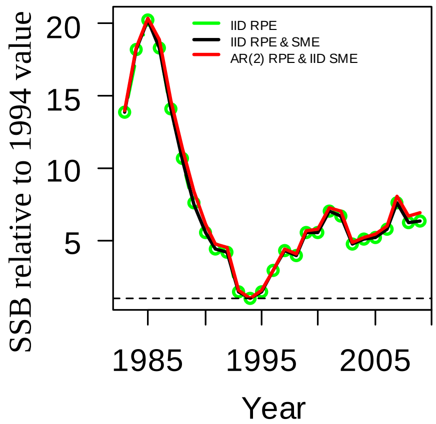


Figure 5.1.9. Trends in SSB relative to 1994 for 3Ps American plaice. Results are from the SURBA+ models with no SSB measurement error ($\sigma_{SME} = 0$; green line), with SSB measurement error ($\sigma_{SME} = .25$; black line), and with AR(2) autocorrelated recruitment process error and SSB measurement errors ($\sigma_{SME} = 0.25$; red line)

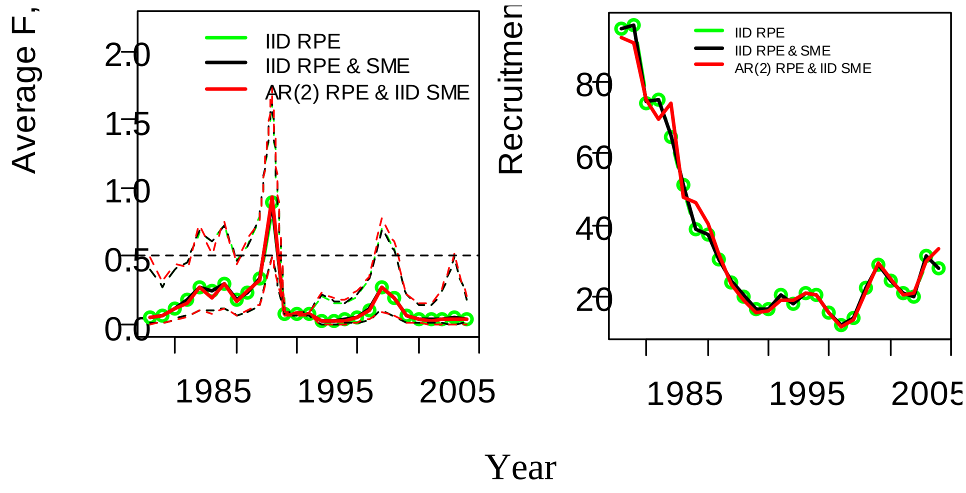


Figure 5.1.10. Average F with 95% confidence intervals (left) and recruitment (right) from three SURBA+ models. See Figure 5.1.9 for other details.

Discussion:

The fishing mortality separability assumptions of SURBA were questioned. These assumptions are been tested, and may not hold, but there is no strong evidence that selectivity has changed.

Why aren't the year effects in the controversial years zero? What other data are being used? The magnitude of the year effect is caused by larger catches in 2009 compared to 2008 for the same cohort (for fully recruited ages), violating fixed catchability assumption. Given no evidence for change in gear, this seems to be a year effect.

Can you see year effects across the multispecies survey in other surveys? This has not been looked at due to different people looking at the different species. In American plaice, there is some similarity in the year effects. The cod was known to aggregate quiet strongly, so hitting these aggregations may cause year effects. Having fished these aggregations down, this probably doesn't happen anymore.

Better modelling of survey data could produce better indices, instead of getting the assessment model to "sort it out".

Why not put autocorrelation into the estimates of year effects? Not sure that there is autocorrelation, and this may be removing information about trends in stock size by attributing them to trends in catchability. Generally, the survey is highly standardized, but trends in time may be possible.

The reason for year effects may be spatial distribution changes, or change in depth, driven by environmental conditions. If there is positive autocorrelation, the confidence intervals on the stock will be underestimated. It would be interesting to fit to residuals after removal of first order autocorrelation, rather than directly to survey.

The importance of being careful when fitting time-series to recruitment residuals was emphasized– similar fixes have been used before to fix average residuals to zero. AR(1) and AR(2) parameters may be correlated strongly.

The effect of autocorrelation in recruitment residuals is that management needs to be more conservative as the probability of multiple years of low recruitment is higher than for IID residuals.

Are the AR parameters well defined given the requirement for a long dataset? The time-series is relatively short, but that is all that is available. Given these data, the options are to either ignore the autocorrelation in SR residuals, or take this approach. Given the short datasets one ought to investigate the effect of misspecifying the AR parameters.

Were function other than AR(1) and AR(2) tried, e.g. Gaussian Bell? No, since that'd normally require more data. The AR(1) is causing longer deviations. It's not clear what the autocorrelation structure should be, but the AR(2) seems to work quite well.

6 Integration of uncertainty (ToR 1e)

6.1 Applicability of various presentations to ToR 1e

In the absence of a contribution specifically tackling this ToR, a discussion was instead held (coordinated by David Miller) on the applicability of various presentations to ToR 1e. All WD referenced are in Annex 5.

Four aspects were considered:

1) Input uncertainty. Uncertainty going into the assessments – identifying it, treating it etc.

2) Handling uncertainty. Treatment of uncertainty by models – can we use outputs to identify excessive uncertainty? Can we design better models that address uncertainty before getting to the point where advice is drafted?

3) Output uncertainty. How certain are we about our uncertainty? Are models masking uncertainty? Can we improve our estimation of it, or do we instead focus on designing management procedures that are robust to it?

4) Management aims uncertainty. Uncertainty in the targets/reference points. Can we improve this?

ToR 1a: Data screening to see if inputs (e.g. survey data) are appropriate

This does not apply directly to the advice, but does apply indirectly via consideration of uncertainty in the inputs.

For the purpose of stock assessments, estimates of uncertainty for cpue measurements need to relate to stock abundance, and inclusion of addition variance over-and-above sampling variance, could help in this regard. The level of aggregation (by haul, by day, etc) is important when deriving estimates of uncertainty.

ToR 1b: Diagnostics to evaluate model fit

Properties of the retrospective and bias indices were analysed and, in particular, the relationship among them (WD 2). Having noticed that some relationship exist between retrospective and bias indices, the potential use of the first index to infer a level of the second one was explored.

Retrospective indices are not a guaranteed way of determining goodness-of-fit (i.e. the lack of a retrospective bias does not guarantee that the model is any closer to the "truth"), yet some limits in retrospective indices could be established indicating unacceptable levels for some bias index.

Potentially, the use of statistical characteristics of abundance indices (in particular varying CV over years) in the stock assessment model, could reduce the retrospective pattern (WD 3) – mainly about reducing uncertainty in outputs, rather than dealing with it for advice.

ToR 1c: Guidance for deciding how complex a stock assessment model needs to be

The move from ASPIC to SS3 for this stock was motivated by the unsatisfactory performance of the ASPIC model, and the desire to make better use of a greater range of data available for the stock (WD 4). How would this impact the uncertainty?

The Bayesian model incorporating the incomplete discard information may be a useful approach with regards to incorporating uncertainty (WD 5). This example essentially removes the assumption that there is no uncertainty by trying to incorporate the discards data. But through the modelling required to incorporate discards, it could increase uncertainty in outputs. This would be an improvement though (better the devil you know than the devil you don't…).

WD 6 considers models that react less to the noise through a statistically relevant handling of the data, thereby reducing the fluctuations in advice.

Here there is a balance between complexity of the model and understanding of trends. Do we need to use all our data every year? Does increasing complexity improve our estimation of the underlying uncertainties?

Simple index based MP

WD 7 considers incorporating uncertainty in the testing of HCRs (developing robust MPs), and simplifying the basis for advice by using simpler methods that allow a more transparent consideration of signals in the data, thereby relying less on prediction and more on observation. Ground test through MSE on a regular but less frequent basis than annual assessments.

ECOKNOWS

An important objective in pooling and handling the knowledge from different sources is to take uncertainty honestly into account. The models suggested will include important knowledge of biological processes and the applied statistical inference methods allow this knowledge to be integrated and updated in stock assessment.

Bayesian inference will form the methodological backbone of the project and will enable realistic estimations of uncertainty. The project aims to improve ways to find generic and understandable biological reference points, and to apply decision analysis and bioeconomic methods to evaluate the validity and utility of improved information.

From ToR 3: guidelines for calculating MSY reference points

Through understanding how uncertainty impacts on our reference points, can we move towards reference points that are more robust to this uncertainty and therefore provide more fixed guidance towards the objectives we aim to achieve? For example (WD 8), if S50 is more sensitive than say Bmsy, would it be more appropriate to use the latter in some way for guidance?

6.2 EU ECOKNOWS project

There has been a request from the EU ECOKNOWS project to present to WGMG what they are doing. The project basically aims to more effectively utilize existing databases and publications, in addition to existing stock-specific knowledge, in order to improve biological knowledge in fish stock assessments (see http://www.ecoknows.eu/). This is an initial introduction to the project, prepared by Samu Mäntyniemi.

Much of the knowledge of biological experts in stock assessment WGs is not reflected in assessment models. The knowledge is probably utilized in an *ad hoc* manner to evaluate whether the results obtained from assessment models are sensible and in trying to understand why the assessment results are what they are. The purpose of ECOKNOWS is to reverse the order so that the assessment models are built from the biological knowledge of the WG members and thereby already include their assessment of what makes biological sense and what does not.

Knowledge from other stocks of related species is not typically formally taken into account in assessments. ECOKNOWS tries to provide conceptual and technical tools to overcome this problem.

The EU 7th framework funded project includes 14 partners and lasts 48 months during 2010–2014. More information can be found from the project webpage: www.ecoknows.eu

ECOKNOWS seeks to develop a generic framework for size-based population dynamics with a biologically plausible life cycle based on ecological knowledge of animal population dynamics. The population dynamics is going to be formulated as a stochastic state-space model with a latent biological structure. The latent model is then linked to observable data by explicitly modelling the processes of collecting the datasets that are going to be available.

This framework should allow for formal ways to account for existing biological knowledge as expert judgement, information accumulated in biological databases and/or estimates from assessments of other stocks of the same or related fish species. Learning from multiple populations and datasets is going to be achieved using hierarchical meta-analysis techniques.

An important objective in pooling and handling the knowledge from different sources is to take uncertainty honestly into account. While this is admittedly going to be difficult, it should lead to more credible assessments. For example, it is highly incredible that anyone could know the rate of natural mortality exactly. It is also unbelievable that a fishery biologist would have absolutely no idea about the values that the natural mortality rate could take. Consistent handling of this type of knowledge can be achieved by using the Bayesian approach to scientific reasoning, where knowledge is measured with probability statements, and probability theory is then used to make updates of the knowledge in the light of new data.

The modelling framework will be used to evaluate the effects of adding new information. The implications of using different types of information on the management reference points will be evaluated and the value of collecting completely new information will be evaluated from the management point of view. The Value of Information (VoI) can be evaluated within the Bayesian decision analysis framework, where the management actions and management objectives are coupled with the system model. Expected utilities of management actions can then be evaluated under uncertainty. The VoI analysis is then the task of assessing the changes in expected utilities and decision rankings under potential new datasets to be collected. This requires that observation models must be built for any new datasets for which the VoI is going to be estimated.

The software development takes place on the R platform, which is used to process modelling inputs and outputs and to act as an interface to JAGS (Just Another Gibbs Sampler) which serves as an engine for Markov chain Monte Carlo (MCMC) simulation. JAGS is a general-purpose MCMC program, which enables fast model development but cannot be expected to deliver optimal efficiency in computation. New MCMC algorithms are going to be developed within ECOKNOWS to speed up the inference in state-space population dynamic models. This development is currently going on in the Matlab environment, but the R interface would eventually be used for the new methods.

At the end of the project a generic model structure should be available with a selection of pre-made submodels for biological processes and fishery and survey observations. These submodels are going to be developed in case studies that are soon going to start applying the current version of the generic model which has been under development for the first year of the project. The case study fisheries include northern hake, northern shrimp, Bothnian sea herring, Atlantic and Baltic salmon and mixed coastal fisheries in Finland and Greece.

The current version of the generic model tracks the binned length distribution over time and treats growth, stock–recruitment and mortality parameters as uncertain variables. The demographic stochasticity is modelled by assuming correlation between survival events of individuals, which leads to an over-dispersed multinomial process for the numbers of fish that survive and get caught by the fishery. The model can take total catches in numbers and length distributions in catch as input data by which the prior distributions put on the model parameters can be updated.

ECOKNOWS relevance to ToR

Data screening -> model choice (ToR 1a):

One of the cornerstones of Bayesian inference is that models by which data are going to be interpreted should not be based on the same data that is going to be interpreted. Instead, the models that are used to interpret data should be made based on prior understanding about the processes that give rise to the data. The purpose is to avoid double use of information. Indeed, in the ideal situation models for data are constructed before any data has been observed. This would make it possible to use VoI analysis to decide what type of data should be collected.

In the context of ECOKNOWS, the assessment models should come from biological knowledge of the stock, and particularly knowledge that is available without looking at the assessment data that is going to be used. This might lead to a situation where multiple models seem possible a priori, and the parameters of each model are highly uncertain. Thus, uncertainty is highest when no data are used, and decreases as the model set becomes conditioned to larger amounts of data.

At least at a first glance this seems to be almost exactly opposite to what is considered in this ToR. Thus, ECOKNOWS is unlikely to help in this area except for providing an alternative way of looking at the necessity and justification of model building based on data exploration. Using the dimension of the data as a guideline for the complexity of the assessment model may force the analyst to imply more knowledge than is actually possessed by the expert WG. A common example is the practice to assume known natural mortality for the sake of identifiability of other parameters. Thus, the uncertainty arising from not knowing M exactly and not being able to estimate it based on assessment data becomes omitted from the analysis and results will look overly precise. Once more data becomes available, the analyst may decide to change assumptions by freeing up a parameter that was previously assumed fixed. This results in an increase of uncertainty compared to adding data and keeping the parameter fixed. Consequently, accumulation of data slowly makes the model more plausible but the uncertainty may seem to stay at about the same level, while one would expect a decrease of uncertainty when more data becomes available.

Model diagnostics (ToR 1b):

ECOKNOWS will develop predictive model checking procedures to examine the predictive goodness-of-fit of Bayesian state-space models. For example, graphical methods to examine the model's ability to predict the length distribution of future catches. The purpose is to be able to qualitatively assess the performance of the model. If uncertainty exists about model structures, ECOKNOWS will try to use Bayesian model averaging to weight the models based on their prior probabilities and their ability to predict data not previously seen by the model.

Information criteria (DIC, BIC, AIC) are not of primary interest in ECOKNOWS. These are suitable methods for statistical data analysis and data compression, where the purpose is to be able to make copies of observed data with as small number of parameters as possible. In ECOKNOWS the primary interest is to make inference about the underlying biological process by using the observed data and prior information. Because the latent process can never be observed, the validity of inference can only evaluated by evaluating the validity of assumptions that are behind the model structure. Being able to predict the future data before using it in the model is still a desirable property, but of secondary importance to the quality of the assumptions made.

Data aggregation (ToR 1c):

ECOKNOWS develops methods to estimate the Value of Information for different kinds of datasets. For example, in the herring case study the value of acoustic surveys and EU data collection catch sampling procedures are going to be evaluated. The same methods can be in principle used to evaluate the VoI of different data aggregation levels. However, in order to do this, it is necessary to build an observation model for the disaggregated data to be able to evaluate the loss of information due to aggregation. But if the observation model for the disaggregated data exists, it might be best to use that one anyway, provided that there are no significant computational costs.

Correlated errors (ToR 1d):

ECOKNOWS case studies will build case specific observation models. Modelling the observation process must include any correlations known to exist in the data collection process. For example, in the herring case the spatial correlation of acoustic survey observations needs to be taken into account. Once the observation model includes the correlation the uncertainty arising from the correlation becomes automatically accounted for.

Integration of uncertainty (ToR 1e):

As discussed above, this is one of the main objectives of ECOKNOWS. Choosing to use the Bayesian approach guarantees theoretical consistency in the integration of uncertainty. However, the Bayesian approach also brings difficult practical challenges to be solved. The main challenges are the derivation of prior probability distributions in practice and the difficult computation of the posterior distributions. ECOKNOWS tackles these issues by developing state-space model structures that would be biologically realistic but yet easily handled in MCMC simulation. This includes reparameterization and development of approximations for the state-space transition equations. New kinds of MCMC samplers that are specific to the state-space models are also under development in the project. Hierarchical meta-analysis tools are going to be built to be usable in an online database environment. FishBase is being used as the test bench for the prototyping. ECOKNOWS case studies will face the problems of deriving prior distributions for the population dynamic parameters. Lessons learned in this process will be summarized in a "Best practices" manual.

Cpue standardization (ToR 2):

Not directly in ECOKNOWS task list, some case studies may need to do this.

MSY reference points (ToR 3):

ECOKNOWS aims to explore how to estimate posterior distributions of MSY reference points. In many of the case studies MSY may appear to be a moving target that fluctuates with environmental variation. Depending on the model structures in case studies, analytical solutions for MSY reference points as a function of population dynamics parameters may or may not be available. However, if the time span for sustainability can be clearly defined, MSY targets can be found out by Monte Carlo integration.

Discussion:

A question was raised regarding the fact that the approach put forward by ECOKNOWS aims to build realistic models, without data availability being a prime consideration. However, such realistic models can be overparameterized with respect to the available data and computationally challenging, requiring in some cases making approximations to the original model structure. Hence, the question was posed as to whether the intended model realism might not be lost due to the required computational approximations. The presenter replied that a key aspect to handle overparameterization is to construct informative and realistic prior distributions. If data are gathered, then the resulting posterior distributions will be even more informative. Additionally, the presenter indicated that computational difficulties should be considered as technical problems, which are expected to lessen in time, and that these challenges should not change the philosophy of building realistic models.

The potential to use Bayesian model averaging to handle uncertainty in model structure was mentioned in the presentation. Bayesian model averaging requires assigning prior probabilities to each model considered and estimating posterior model probabilities, which are then used as weights for the model averaging. Again, the question of whether this might not be computationally too challenging (requiring very complex MCMC algorithms) was raised. The presenter replied that certain simplified forms of model uncertainty could be handled reasonably easily. For example, when considering two different possibilities for a stock–recruitment relationship (such as Beverton–Holt and Ricker) one could simply introduce an auxiliary variable with Bernoulli prior distribution, giving equal prior probability to each of the two forms of stock–recruitment relationship. The posterior distribution of this auxiliary variable gives the posterior probability of each of the two stock–recruitment relationships and model averaging is directly incorporated in the model results.

7 Commercial cpue standardization (ToR 2)

7.1 Cpue standardization— methodologies and practices

(compiled by Sam Subbey)

Background

Fisheries-dependent data (e.g. scientific surveys) provide a basis for calculating indices of abundance. However, collecting fisheries independent data are usually costly, in financial terms, and with respect to time and human and material resources required.

On the other hand, Fisheries-dependent data (e.g. catch and effort data from commercial and/or recreational fishers) can be summarized into catch rates and/or catch-per-unit-effort (cpue).

A common definition of cpue is a ratio of the total catch and the corresponding fishing effort over a specific spatial scale and time. The cpue is commonly used as an index of fish stock abundance, implying that a proportional change in cpue is expected to represent a matching proportionate change in the stock size. A usual underlying assumption is that within a defined spatial scale, the distribution of fishing effort with respect to targeted fish is random.

The resolution of the spatial scale however, will have an influence on estimates of cpue. Further, nominal cpue values seldom reflect changes in abundance over the whole exploitation history of spatial range of the stock. This is because cpue indices can be affected by exogenous factors which bear no relationship with changes in stock abundance, e.g. the vagaries of the weather, fuel prices, etc.

The aim of the standardization process is to remove the effects of spatial and temporal changes in extraneous factors on catch rates so that changes in the standardized index reflect changes in abundance only (Maunder and Punt, 2004).

Standardization – models

The accepted approach is to model the expected value of the cpue (or a function of the expected value) as dependent on area, seasonal, and other interactive factors. In quite a general way:

Expectations of [function of cpue] = (Intercept) + (Year) + (Area) + (Season) + (environmental factors, fishing gears, operating devices, etc.) + … + (Interactions),

where (Year): effect of year, (Area): effect of area; (Season): effect of month/quarter; (environmental factors, fishing gears, operating devices, etc.): effect of environmental factors such as sea surface temperature, fishing gears, operating devices, etc. (Interactions): two way interactions.

The most common methods for cpue standardization involve fitting statistical models to catch and effort data. These models include Generalized Linear Models (GLMs), Generalized Additive Models (GAMs) and Generalized Linear Mixed Models (GLMMs). Of the three, GLMs are the most common methods in use for CPU standardization. A brief description of the models is presented below.

GLMs

The relationship between some function of the expected value of the response variable and the explanatory variable is linear: $g(\mu_i) = x_i^T\beta$, where g is a monotonic and differentiable link function, x_i and β are vectors of explanatory variables and parameters, respectively, and $\mu_i = E(Y_i)$, where Y represents random variables.

The GLM modelling approach requires the choice of:

i. Response variable and its sampling distribution from an exponential family of distributions (normal, exponential, Poisson, binomial or gamma).

ii. Appropriate link function, consistent with choice of distribution. For example, the Logit function $g(k) = \ln\left(\frac{k}{1-k}\right)$ is appropriate for a binomial distribution.

iii. Explanatory variables (e.g. Year).

GAMs

GAMs are extensions of GLMs where the linear predictor in GLMs are replaced by an additive predictor by defining $g(\mu_i) = \mu + \sum_{j=1}^{p} f_j\,(x_i)$, where f is a smooth function, such as a spline or loess smoother.

GLMMs

GLMMs are based on extending GLMs to include random effects, random coefficients and covariance patterns. The link function in GLM, $g(\mu) = X\beta$ is redefined for GLMMs as $g(\boldsymbol{\mu}) = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\alpha}$, where $\mathbf{X}$ is the design matrix for fixed effects, $\mathbf{Z}$ defines the design matrix for random effects, $\boldsymbol{\beta}$ is a vector of fixed effect parameters and $\boldsymbol{\alpha}$ is a vector of random effect parameters, assumed to follow a normal distribution.

Challenges

The main challenges include selecting explanatory variables, model choice and assumptions of error structure, data selection and dealing with zero catches.

Selecting explanatory variables

In the literature, the explanatory variables include those of time (year, month, time of the day), area, type of trawl gear, vessel characteristics (size, length, etc). Cross validation – determination of optimal model parameters, using a subset of the data (training set) and prediction of the rest of the data (test set) using the optimized model—is an approach for selecting explanatory variables.

Caveats: Collinear variables and risk of over/under-parameterization.

Model choice and assumptions

Standard hypothesis testing methods (F-tests, likelihood ratio test, etc) are directly applicable to nested models. Information-theoretic methods (e.g. AIC and BIC) are applicable to non-nested models.

Caveats: The modelling is based on voluminous datasets, hence the best model may still be a highly parameterized model chosen based on AIC or BIC. Testing for validity of model assumptions is largely neglected (e.g. residuals from log-linear regression are normally distributed).

Data quality

One important issue not addressed in the literature is the effect of data aggregation. Whether cpue data are aggregated on hauls/monthly or daily, will have effect on accuracy and precision. For instance, monthly averages will minimize variability while individual haul cpues will have larger variability and hence variance. The literature, however, deals with issues of zero catches under positive effort data scenarios.

Dealing with zero catches

Quite generally, cpue data are non-negative and severely skewed right, suggesting gamma or lognormal models. However, often no fish are caught, which produces exact zeros in the data. Zero catches are common for less abundant and bycatch species, where non-zero effort is registered to correspond to zero catch. These exact zeros cannot be ignored and contain important information. However, zero catches present computational difficulties, e.g. for log-linear (and gamma) models, since the natural logarithm of zero is undefined.

Appropriate methods for handling zero catches in the literature include adopting:

1) Ad hoc approach where a small constant is added to all response variables (such as cpue), e.g. E[log(CPUE + constant)] = (Intercept) + (Year) + (Area) +(Season) + (EMT) + … + (Interactions).

2) Use of the Catch-Poisson or Catch-Negative-Binomial (NB) regression models (i.e. Catch model with Poisson/negative binomial error, GLM-type) and extensions using the Tweedie distribution.

3) Zero-inflated models.

4) Use of the delta-type two-step model (e.g. Delta-lognormal model).

The Tweedie, zero-inflated and delta-type models are described briefly below.

The Tweedie distribution model

The Tweedie distribution model, $f(y|\mu, \sigma^2, p)$, is a 3-parameter model defined by:

$$f(y|\mu, \sigma^2, p) = a(y|\sigma^2, p)\mathrm{e}^{\left\{-\frac{1}{2\sigma^2}\mathrm{d}(y|\mu,p)\right\}}$$

Where σ^2 and μ are location and diffusion parameters, respectively, p is the power parameter, and $\mathrm{d}(y|\mu, p)$ is referred to as the unit deviance. This power-parameter (p) can be defined as an arbitrary real number except for $0 < p < 1$. The Tweedie model can express the Poisson, Gamma and inverse Gaussian distributions if the power-parameter (p) is 1, 2, and 3, respectively.

In the literature, application of the Tweedie model involves a 2-step approach:

i) Estimate the power parameter (p) by maximizing the profile log-likelihood across the grid values of (p) in the range of $1 < p < 2$.

ii) Estimate the regression coefficients (in e.g. GLMs) fixing the value of p in the estimate obtained in the step i.

Zero-inflated models

Zero-inflated count models provide a way of modelling the excess zeros in addition to allowing for over-dispersion. In particular, for each observation, there are two possible data generation processes; the result of a Bernoulli trial determines which process is used. For observation i, Process 1 is chosen with probability w_i and Process 2 with probability$(1 - w_i)$. Process 1 generates only zero counts, whereas Process 2, $f(y)$, generates counts from either a Poisson or a negative binomial model.

The general expression is

$$Pr(Y = y) = \begin{cases} w + (1 - w)f(0), & y = 0, \\ (1 - w)f(y), & otherwise, \end{cases}$$

where w is the probability that an observation comes from the degenerate component. The parameters to be modelled as functions of the explanatory variables are the probability of a zero observation, w, and the mean of the second distribution defined by $f(y)$. The two commonly used zero-inflated distributions are the zero-inflated Poisson (ZIP) and the zero-inflated negative binomial (ZINB).

The proportion of zeros in the Poisson and negative binomial distribution is related to the distribution for the non-zero values (i.e. for a given distribution of non-zero observations there is only a single possible proportion of zeros). However, if the processes that lead to zero observations are not the same as those that lead to non-zero catches (e.g. gear malfunction, whether the species under consideration is being targeted), zero-inflated distributions may be more appropriate.

Delta approaches

The delta approach models the probability of obtaining a zero catch and the catch rate, given that the catch is non-zero, separately.

The general expression is

$$Pr(Y = y) = \begin{cases} w, & y = 0, \\ (1 - w)f(y), & otherwise, \end{cases}$$

Where w is the probability of a zero observation (not the probability of an extra zero, as in the zero-inflated approach). The probability of obtaining a zero observation is usually modelled using the binomial distribution, while the lognormal distribution has often been used to model the catch rate given that it is non-zero.

This section contains extracts and summaries from several papers. Key among these papers are Maunder and Punt (2004), Shono (2008a, b), Tweedie (1984), and Venables and Dichmont (2004).

7.2 GLM standardization using the Tweedie distribution

The techniques applied for cpue standardization are based mainly on the regression model, and contemporary approaches use Generalized Linear Models (GLM). GLM models are based on exponential family of distribution that includes the Normal, Poisson, Binomial, Gamma and Inverse Gaussian distributions. Each of these distributions is completely specified by its mean and variance. The variance of the response variable is a function of its mean.

The Tweedie distribution also belongs to the exponential family but requires three parameters: the variance function is proportional to the power of the mean. This power parameter (p) determines the family of different distributions: normal (p=0). Poisson (p=1), Gamma (p=2), Inverse Gaussian (p=3), compound Poisson-Gamma (1<p<2). The main properties of the Tweedie distribution are the ability to handle very high variability, highly skewed distributions and exact zero values.

The GLM with the Tweedie distribution can be used in the same way as other GLM models but it is necessary to do a set of runs with different values of parameter p to find the minimum value of the extended deviance profile.

GLM models with mixed effects contain not only fixed effects coefficients but also coefficients describing random effects. Diagnostics of the models include ANOVA estimates, Pearson residuals as functions of fitted values, deviance residuals as functions of fitted values, and Q-Q plots. Results can be presented using plots with fixed effects time-series and random effects accompanied by estimates of bias.

S-Plus and R software packages have appropriate functions (reglm, tweedie) for these type of models.

Examples contain results of the calculation with both GLM and GLMM models, including calls to the function, diagnostic results and time-series of the cpue estimates. Applications of GLMM models to obtain cpue time-series that describe indices of abundance in different parts of the fishing area, important for fitting spatially structured stock assessment models, was also shown (presentation only – no WD provided).

Discussion:

Generalized linear models (GLM) have become quite common tools for cpue standardization. The presentation reviewed GLMs, which are based on distributions from the exponential family. Distributions from the exponential family have 2 parameters, one of which determines the mean whereas the second one controls dispersion. Each distribution in the exponential family is characterized by a function that relates the variance of the distribution to its mean. The presentation highlighted the fact that Tweedie distributions, which are 3-parameter distributions, belong to the exponential family for each given value of the so-called variance power parameter "p". When p=0, the Normal distribution is obtained, p=1 corresponds to Poisson distribution, p=2 to the Gamma distribution and p=3 to the Inverse Gaussian distribution. For each value of p between 1 and 2, the Tweedie distribution is a mixture of a point mass at zero and a continuous distribution on the positive real line. This can be useful for modelling cpue data, which may contain a substantial amount of zeros. To select an appropriate value of p, a separate GLM may be fitted over a grid of different values of p and then examine the likelihood profile for p, selecting the value of p for which the likelihood profile is minimized. Random effects via Generalized Linear Mixed Models were also considered in the presentation as well as possible interactions between factors, possibly also involving "year".

It was agreed that the assumption of a random effects model together with fixed effects is more realistic, particularly if interactions are considered. Results were expected to be differ depending on whether interactions are considered or not, but both solutions should be explored. The no interactions model will allow one to know the deviation distribution and to check if it is the same for any classification criterion. Year, month and area are primary candidates when searching for interactions.

Interaction terms with month can be handled through averaging over months, with area by multiplying by the size of the area, but what to use as an index of abundance when there are significant interaction terms that include year, when an annual index is required?

GLMs give unrealistically low estimates of CVs on, for example, the standardized cpue, because they assume (incorrectly) that the input data are independent. Random effects models (treating interactions with year as the random effects) is one way to deal with this, though there are issues that warrant further investigation. A simple approach would be to do a jackknife with year as the sampling unit as a way of taking the non-independence into account.

The problem of many zeros can be reduced through aggregation. In that context of aggregation, it is more robust and less variable to use sum of catch divided by the sum of effort over the entire aggregation period as opposed to e.g. calculating daily cpue and averaging over the aggregation period (e.g. a month).

The method presented allows using haul by haul catch data, or grouping them by day or by month. Haul by haul catch data are quite skew in their distribution, having lots of zeros in some cases, but the Tweedie distribution fits adequately such situations. Grouping data by month would result in a less skew distribution but will still be well explained by a Tweedie distribution with a different value of the variance power parameter p. In the work presented, the interannual dynamics of standardized cpue based on data aggregated by month was very similar to that for haul-by-haul data or other scales of fishery data aggregations. The aggregation of data allows an improvement in diagnostics and results from cpue standardization based on GLMM with Tweedie's distribution.

8 MSY in a stochastic environment (ToR 3)

To some extent, this ToR has been explored by other groups within ICES. For example, WKFRAME-2010 (ICES, 2010a) provides comprehensive guidelines for calculating MSY reference points, which was subsequently updated in the WKFRAME-2 report in 2011 (ICES, 2011). Furthermore, the WKFRAME-2010 report points to guidelines in the SGMAS 2008 report (chapter 5.2 of ICES 2008b) with recommendations for how to use a stochastic simulation model that could be used for estimation of MSY reference points, taking risk to recruitment impairment into consideration and incorporating density-dependent process into analyses. Nevertheless, this topic was also explored during this meeting.

8.1 Correcting for measurement error bias when fitting stock-recruit models and estimating MSY reference points

Dealing with stock size measurement errors when fitting stock–recruit models, and in particular how these affects MSY reference points, can be challenging. This was a particular concern for WGs in 2010 that were expected to provide MSY reference points, but were faced with how to account for sometimes poor fitting stock–recruit relationships.

Consideration is given to the development of a purely statistical approach to address the problem of the impact of varying productivity on MSY reference points (Annex 5, WD 8). The few papers in the literature that have addressed the problem in some way have taken a more multispecies/ecosystem approach in which they try to account for global warming and other such issues. However, the "state-of-art" for many stocks in regions including Europe are not at the stage where these multispecies approaches can be used. Nevertheless, there is a demand to address the problem, and the proposed development of the statistical approach is seen as an interim solution.

Discussion:

This study demonstrates that in case of substantial error in the spawning-stock biomass, the approach involving the estimation of reference points (RPs) based on the S50% parameter derived from the Beverton–Holt model may be seriously biased. It should be decided if such a sensitive parameter could be used as a basis for RP estimation. The use of local influence diagnostics did not help with the estimation of measurement error, and additional information is needed. Further to the results reported, it would be useful to check how assumptions about error structures in the data may influence the results obtained.

The assumption that an autocorrelated SSB implies that the errors associated with it are also autocorrelated is not always valid. If the SSB is for instance, estimated from indices of abundance, it may well be correlated although the errors (linked to the abundance indices) may be random.

Subsequent work:

Guidelines for calculating MSY reference points in a varying and stochastic environment

Conceptually, calculation of MSY reference points (RPs) involves evaluating long-term stock projections in which fishing mortality is varied to find the level (Fmsy) that maximizes long-term yield. MSY is the maximized yield and Bmsy is the equilibrium stock size that gives MSY. If the projection is deterministic then the calculation of MSY RPs is also deterministic. In this context some theory has been developed to simplify MSY calculations (Sissenwine and Shepherd, 1987). In the traditional MSY calculations, all population processes are assumed to be constant; that is, the age-based values of natural mortality, maturity, weight, and fishery selectivity in the spawner-per-recruit relationship are held constant in the long-term stock projections, as is the recruit-per-spawner functional relationship with SSB. Estimation error in these population processes contributes to uncertainty, and some bias, in MSY RPs.

If the population processes themselves are variable, then MSY RPs will also vary. For example, if natural mortality (M) changes in future, as a function of predators or other factors, then this will affect MSY RPs. If the population processes that vary have, or are expected to achieve, a stationary distribution then there may also be stationary distributions for MSY RPs that are useful for fisheries management, although the RPs will be random and this should be accounted for in management decisions.

In a stochastic environment, harvesting according to the deterministic MSY rule is an under-optimized strategy and can lead to strong decreases in stock size (Bousquet *et al.,* 2008). These authors showed for the Schaefer surplus production model with a particular type of bounded process error that the stochastic mean values for MSY, Bmsy and Fmsy, were less than the deterministic results. They concluded that the deterministic Fmsy is incompatible with the assumption of equilibrium: on average, one cannot hope to harvest more than the stochastic MSY. Constant harvesting at the deterministic Fmsy would eventually lead to stock extinction.

Some preliminary investigations are presented on the impact of process error in the stock–recruit relationship and measurement error in SSB on MSY RP's. Analyses are based on the SURBA+ model results for 3Ps American plaice (Annex 5, WD 8); however, the choice of assessment model is not relevant beyond providing the stock–recruit relationship and values for process error variances. Stochastic simulations are used to find the fishing mortality rate (Fmsy) that optimizes long-term equilibrium expected catch and produce equilibrium distributions for biomass and catch at F = Fmsy.

Independent recruitment process error

Equilibrium distributions for MSY RP's were derived using stochastic projections in which recruitment was derived using

$$R(S) = \frac{\alpha S e^{\varepsilon_{RPE}}}{\beta + S},\ \varepsilon_{RPE} \sim N(0, \sigma_{RPE}^2)$$

The stock–recruit parameters and σ_{RPE} were estimated using SURBA+, with the same assumption for recruitment process errors (RPE's). The estimates are $\hat{\alpha} = 27.832$, $\hat{\beta} = 1.698$, and $\hat{\sigma}_{RPE} = 0.315$. The stock size estimates are in kg/tow, and recruits are in number per tow. Ten thousand projections, each for a 100 years, were conducted. The process errors were bias corrected so that E{exp(ε_{RPE})} = 1. The multiplicative errors were $\exp(\varepsilon_{RPE} - \sigma_{RPE}^2 / 2)$.

Equilibrium mean yield as a function of F was similar to the deterministic results (Figure 8.1.1), which were derived using the same stock–recruit parameters but with $\sigma_{RPE} = 0$. Contrary to the conclusions in (Bousquet *et al.,* 2008), the stochastic Fmsy was slightly greater than the deterministic result. The resulting equilibrium biomass and catch when Fmsy = 0.232 are shown in Figure 8.1.2. The mean equilibrium SSB (16.5 kg/tow) is slightly lower than the deterministic result (17.08 kg/tow). Stochastic and deterministic MSY catches were nearly identical. Note that the means and percentiles are stable after about 30 years, indicating that an equilibrium distribution is achieved when F = 0.232.

Independent recruitment process error and SSB measurement error

Equilibrium distributions for MSY RP's were derived using stochastic projections in which recruitment was derived using

$$R(S) = \frac{\alpha S e^{\varepsilon_{SME}}}{\beta + S e^{\varepsilon_{SME}}} e^{\varepsilon_{RPE}},\ \varepsilon_{RPE} \sim N(0, \sigma_{RPE}^2)\ and\ \varepsilon_{SME} \sim N(0, \sigma_{SME}^2)$$

The stock–recruit parameters and σ_{RPE} were estimated using SURBA+, with the same assumption for recruitment process errors (RPE's). Because σ_{RPE} and σ_{SME} are confounded, σ_{SME} was fixed at 0.25 for illustration purposes. The estimates are $\hat{\alpha} = 28.661$, $\hat{\beta} = 1.960$, and $\hat{\sigma}_{RPE} = 0.309$. Both the lognormal recruitment process errors and SSB measurement errors were bias corrected (so that E(ε)=1) in the projections.

Equilibrium mean yield as a function of F was similar to the deterministic results (Figure 8.1.3). The stochastic Fmsy was slightly greater than the deterministic result. The resulting equilibrium biomass and catch when Fmsy = 0.226 are shown in Figure 8.1.4. The mean equilibrium SSB (17.2 kg/tow) and catch (4.67 kg/tow) are slightly lower than the deterministic results (17.79 kg/tow, 4.71 kg/tow respectively).

Autocorrelated AR(2) stock-recruit residuals

This is an extension of the model in the previous section in which the ε_{RPE} 's are AR(2) autocorrelated. The Beverton–Holt parameter estimates are $\hat{\alpha} = 28.596$, $\hat{\beta} = 1.828$. The estimated AR(2) autocorrelation parameters are 0.902 (lag 1) and -0.541 (lag 2) with $\hat{\sigma}_{RPE} = 0.238$. Both the recruitment process errors and SSB measurement errors were bias corrected in the projections.

Equilibrium mean yield as a function of F was similar to the deterministic results (Figure 8.1.5). The stochastic Fmsy was slightly greater than the deterministic result. The resulting equilibrium biomass and catch when Fmsy = 0.228 are shown in Figure 8.1.6. The mean equilibrium SSB (17.0 kg/tow) and catch (4.70 kg/tow) are slightly lower than the deterministic results (17.63 kg/tow, 4.74 kg/tow respectively).

Sensitivity runs

The lognormal process and measurement errors were rescaled to have means of one. If the errors are not rescaled then the stochastic mean MSY catch and Bmsy can be greater than the deterministic results. This is because the means of the process and measurement errors are greater than one. This was observed for the AR(2) analysis. However, the Fmsy values changed little. When errors were not rescaled to have mean one, Fmsy = 0.229 whereas when the errors were rescaled Fmsy=0.228.

The effect of doubling the recruitment process error variance was also investigated. The changes in mean Fmsy were negligible for the scenario with independent recruitment process errors and SSB measurement errors. Mean SSBmsy and MSY catch decreased slightly, from 17.2 kg/tow and 4.67 kg/tow to 17.1 kg/tow and 4.65 kg/tow, respectively. The largest differences were in the quantiles of the stochastic distributions for SSBmsy and MSY. They were wider when the recruitment process error variance was doubled, as expected. A similar pattern was observed for the AR(2) recruitment process error scenario.

Summary

Stochastic Fmsy's were usually slightly greater than deterministic Fmsy's. This is not consistent with Bousquet *et al.*, 2008. Stochastic mean MSY catches were similar to the deterministic results, which is also not consistent with Bousquet *et al.*, 2008. However, Stochastic mean SSBmsy was usually slightly lower than the deterministic SSBmsy.

This is consistent with Bousquet *et al.* (2008), but the differences in stochastic and deterministic results were usually small.

The amount of process error had little effect on mean MSY reference points, but did affect quantiles. Also, constraining process errors to have mean one made a difference.

These results are preliminary and more research is required.

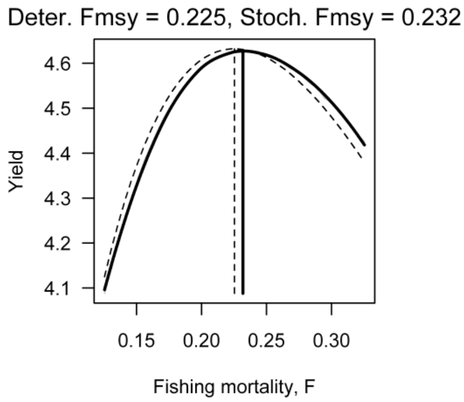


Figure 8.1.1. Equilibrium mean yield as a function of F, when there is independent process error in the recruitment derived from a Beverton–Holt stock–recruit model (solid line), and when there is no process error (dashed line).

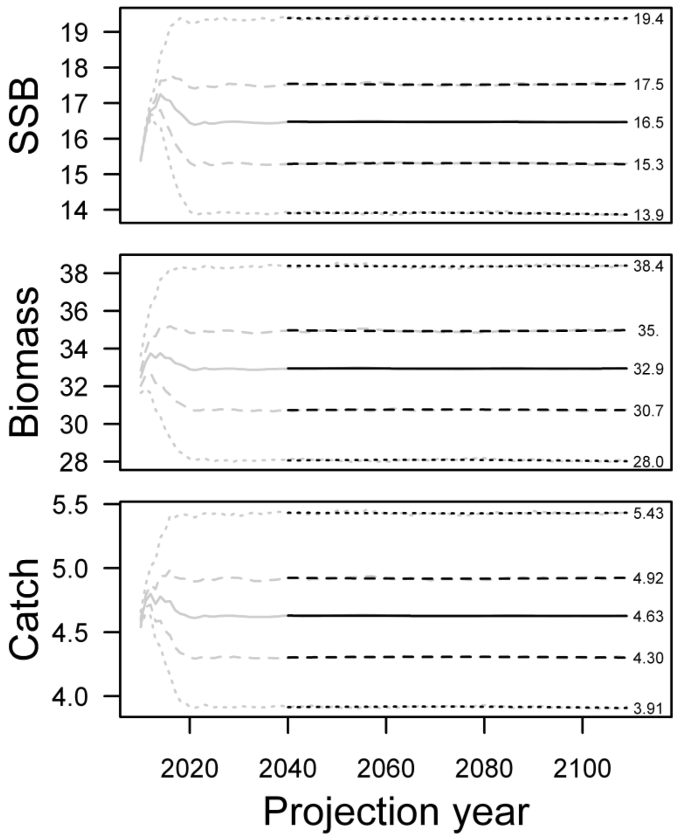


Figure 8.1.2. Equilibrium results for SSB (top panel), total biomass (middle panel), and catch (bottom panel), for projections based on Fmsy = 0.232 (see Figure 8.1.1). The solid lines indicate the means of the stochastic distributions, the dashed lines indicate the 25th and 75th percentiles, and the dotted lines indicate the 5th and 95th percentiles. Light grey lines are simulation results, and the black lines are results from a loess smoother. They are virtually identical after about 30 years. The deterministic SSBmsy is 17.08, and the deterministic MSY catch is 4.63. All results are in kg/tow.

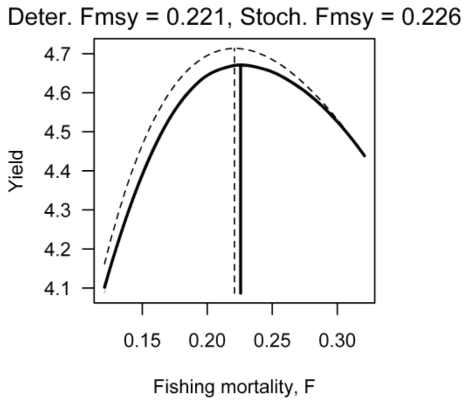


Figure 8.1.3. Equilibrium mean yield as a function of F, when there is independent process error in the recruitment derived from a Beverton–Holt stock–recruit model with SSB measurement error (solid line), and when there is no process or measurement error (dashed line).

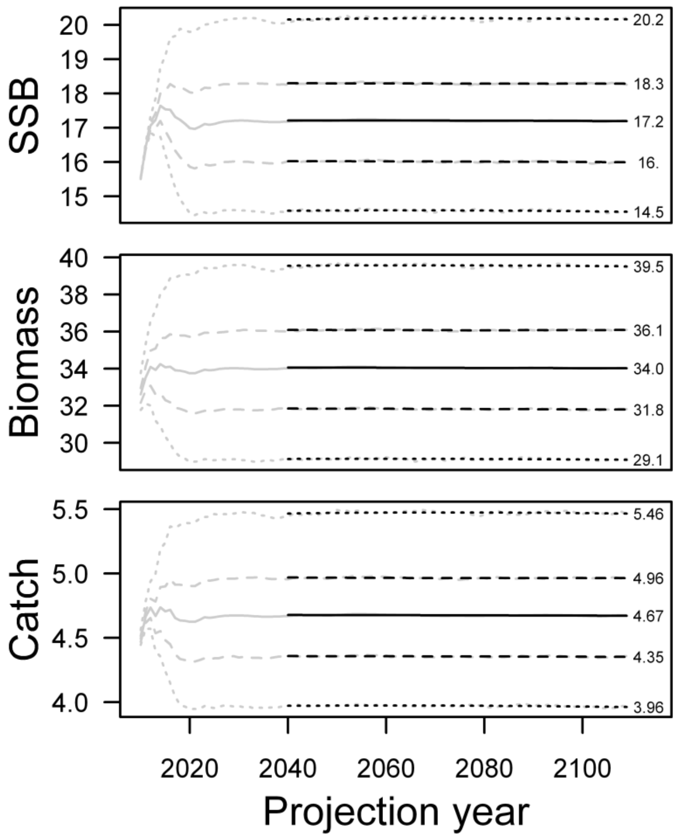


Figure 8.1.4. Equilibrium results for SSB (top panel), total biomass (middle panel), and catch (bottom panel), for projections based on Fmsy = 0.226 (see Figure 8.1.3). The solid lines indicate the means of the stochastic distributions, the dashed lines indicate the 25th and 75th percentiles, and the dotted lines indicate the 5th and 95th percentiles. Light grey lines are simulation results, and the black lines are results from a loess smoother. They are virtually identical after about 30 years. The deterministic SSBmsy is 17.79, and the deterministic MSY catch is 4.71. All results are in kg/tow.

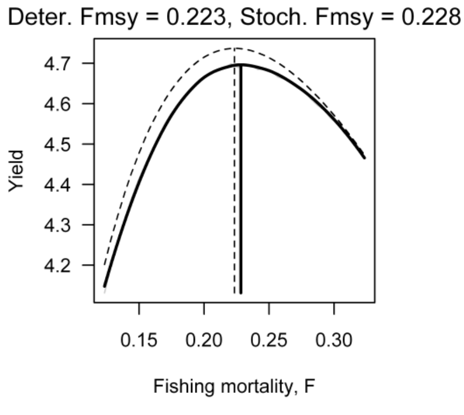


Figure 8.1.5. Equilibrium mean yield as a function of F, when there is AR(2) process error in the recruitment derived from a Beverton–Holt stock–recruit model with SSB measurement error (solid line), and when there is no process or measurement error (dashed line).

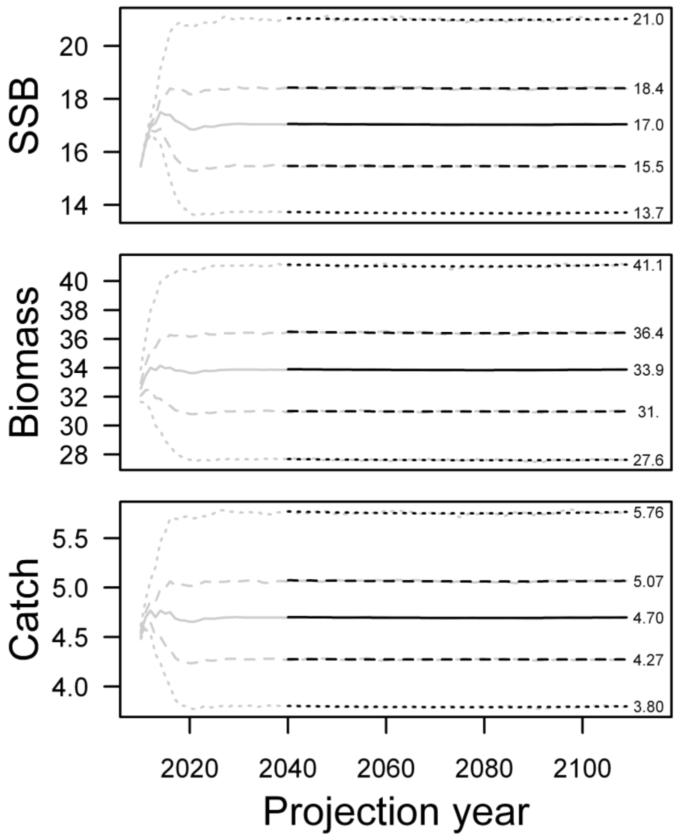


Figure 8.1.6. Equilibrium results for SSB (top panel), total biomass (middle panel), and catch (bottom panel), for projections based on Fmsy = 0.228 (see Figure 8.1.5). The solid lines indicate the means of the stochastic distributions, the dashed lines indicate the 25th and 75th percentiles, and the dotted lines indicate the 5th and 95th percentiles. Light grey lines are simulation results, and the black lines are results from a loess smoother. They are virtually identical after about 30 years. The deterministic SSBmsy is 17.63, and the deterministic MSY catch is 4.74. All results are in kg/tow.

Discussion following subsequent work

Bousquet *et al.* (2008) were critical of people who ignore process error, but the results of this work don't seem to indicate that it's important. However, in a model that incorporates process error, the number of individuals in the population may simply mean that process error CVs are very small – could this be what is leading to the result of this study? Although Surba+ is based on numbers-at-age, it is only used for deriving stock–recruit pairs, and is not used in the projections; furthermore, Bousquet *et al.* used biomass, so process error may well have had a greater impact.

Although the ICES WKFRAME and WKFRAME-2 working groups considered uncertainty in stock–recruit models, they did not make progress in handling environmental variability such as regime changes, and their effect on the estimation of reference points. There may also be other sources of process error to consider.

9 The SISAM Initiative (ToR 4)

9.1 The Strategic Initiative for Stock Assessment Methods

The ICES Strategic Initiative for Stock Assessment Methods (SISAM – chaired by Steve Cadrin and Mark Dickey-Collas) is designed to assure that ICES scientists can apply the best methods when developing management advice. Other Regional Fisheries Management Organizations (RFMOs) and national fishery organizations have a similar goal, so success of SISAM will have benefits for the entire international fishery science community. SISAM will contribute to the improved application of assessment methods, but it must be recognized that "best methods" is not a static definition. Rather, the set of available methods will continue to evolve and improve in response to lessons learned in their current applications. SISAM needs to do more than define the current state-of-the-science; it should help chart the future course of this scientific enterprise. Long-term success in application of the best methods is an iterative, multi-step process. These steps should involve:

1) identification of the current set of available methods;
2) guidance in the selection of the most appropriate methods for a particular application;
3) education and access to expert information regarding method usage;
4) encouragement for further testing and development of methods to more closely align with particular management needs and to take advantage of advances in statistical theory, computing power, and new knowledge.

SISAM can contribute to this process by directly advancing steps 1 and 2 and serving as a valuable catalyst for steps 3 and 4. SISAM proposes to accomplish this by producing a technical report (details below), sponsoring an international symposium on fishery assessment methods (to be held towards the beginning of 2013), and publishing key papers from the symposium in a scientific journal. SISAM will seek to encompass approaches that range from quantitative procedures applicable in data-poor situations, through tactical assessment approaches that typify assessment advice today, to multispecies and environmentally linked models that are at the forefront of research today. Within this range, the principal focus will be on the tactical assessment approaches, with briefer consideration to the data-poor and advanced model categories.

Technical Report
The proposed technical report will combine the developed model categorization scheme (the current draft on which WGMG was requested to comment is presented below) and an overview of recent model usage by a wide range of RFMOs and national organizations. It will provide a structured organization of these models to guide ICES Working Groups in their search for appropriate models for each situation. This will be prepared prior to the 2013 symposium, to act as a resource to guide discussion and stimulate the workshops that shall take place during the symposium. It is proposed that later, the report will be published as an ICES Cooperative Research Report (CRR), which will also include the results from the symposium workshops.

The ICES WKADSAM workshop in 2010 started the process of identification of available methods by bringing together ICES and international assessment experts to describe state-of-the-science assessment models. The SISAM effort will build upon this foundation of model descriptions provided by WKADSAM and through further discussions with the ICES WGMG. SISAM proposes to reach out to RFMOs and national fishery organizations to request information on the methods used to conduct assessments over the past 5 years. This request will need to involve some degree of information about data used because today's generalized integrated analysis models can be applied across a wide range of data types, so information on data used is valuable to refine the information on model usage. Because many methods are essentially similar and differ only in name and details of the particular application, SISAM with WGMG will also develop a categorization system for fishery assessment models. This system will allow for clear delineation of major categories of models, and identification of the models available within each category. The draft report with the summary of model usage and the categorization system would be prepared for availability by the time of the proposed symposium. The final categorization scheme will be agreed by the end of 2011.

Initial work on the categories for classification of models along the "age" axis (presented to WGMG for comment)

0. *Techniques for when the only available information is catch data – this will not be a main focus of SISAM but a summary of available techniques will be provided*
    a. *Structure – some use basic biomass dynamics*
    b. *Min data – catch*
    c. *Typical data – catch, some expert opinion on stock depletion or F*
    d. *Example –Depletion Based-Stock Reduction Analysis (DB-SRA; Dick and MacCall)*
1. *Time-series models*
    a. *Structure – none or minimal assumptions, just examining catch and/or index as time-series*
    b. *Min data – catch or abundance index time-series*
    c. *Typical data – catch and abundance index*
    d. *Example – AIM (US Toolbox), empirical management procedures*
2. *Dynamic Surplus production models*
    a. *Structure – aggregate biomass dynamics controlled by a small number of parameters: typically just K (carrying capacity), r (intrinsic growth rate), initial population biomass and a catchability coefficient related to fishing mortality.*
    b. *Min data - catch and one relative abundance index.*
    c. *Typical data – catch and one or several abundance indices*
    d. *Example – Dynamic Schaefer model, ASPIC*
3. *Delay-difference models:*
    a. *Structure similar to surplus production but with at least two life stages, one typically for fish before recruitment to the fishable pool of the stock, and with some somatic growth relationship and fishing mortality included in the population dynamics*
    b. *Min data – catch, abundance index, inputs for body growth function and M*
    c. *Typical data – catch, recruitment index, recruited (adult) index*
    d. *Examples – Deriso model, CSA, various others involve approaches to dealing with process error and/or state-space formulations*
4. *Age-structured production models*
    a. *Structure – full age structure, use a deterministic spawner-recruitment relationship (which replaces the r and K of the dynamic surplus production models);*
    b. *Min data – catch, fishery selection-at-age, abundance index with specified selection pattern at age, M and body wt-at-age, and maturity-at-age.*
    c. *Typical data – min data plus additional abundance indices*
    d. *Examples – Age-Structured Production Model (ASPM)*

5. *Stochastic stock production models*
    a. *Structure – Same dynamics as age-structured production model but allowing for recruitment to be stochastic rather than deterministic.*
    b. *Min data – as for the age-structured production model,*
    c. *Typical data – catch, abundance index, recruitment index*
    d. *Examples – Walters and Martell model; Porch's Catch-Free model, DB-SRA (Dick and MacCall)*
6. *Integrated analysis models used with length data but no age data:*
    a. *Structure – Population dynamics are age and/or length structured, and incorporate natural mortality, growth, recruitment (which may or not be based on a stock–recruitment relationship with or without deviations), and fishing mortality-at-age and/or length. Some implementations allow treatment of landings and discards.*
    b. *Min data – catch, abundance index, length composition data (some missing data allowed).*
    c. *Typical data – catch, abundance index, length composition data. Some implementations allow the catch data to be separated into landings and discards.*
    d. *Examples – MULTIFAN, simplified configurations of SS and CASAL, SCALE*
7. *Statistical catch-at-age models:*
    a. *Structure – Age-structured population dynamics incorporating natural mortality, recruitment deviations (but most models do not employ internal spawner-recruitment relationships and treat recruitments as free parameters), and fishing mortality (the fishery selection-at-age may be constant or change over time according to some constraints,) ; some implementations have a specialized approach to deal with discarded catch separately from landings.*
    b. *Min data – catch, statistical sample of catch age composition, abundance index (some missing data allowed). Some implementations allow the catch data to be separated into landings and discards.*
    c. *Typical data – catch, statistical sample of catch and abundance index age compositions, abundance index*
    d. *Examples – ASAP, AMAK, many custom ADMB coded applications*
8. *VPA-based approaches:*
    a. *Structure – Population abundance at age directly calculated from catch-at-age (treated as known and without error in every time-step) and M, starting from the latest year and oldest true age for each cohort. Often incorporate fits to age-specific abundance indices.*
    b. *Min data – complete, high quality catch-at-age for every time-step and one abundance index for tuning*
    c. *Typical data – min data and several age-specific tuning indices*
    d. *Examples – XSA, ADAPT, VPA2BOX*
9. *Integrated analysis models:*
    a. *Structure – Basically same population dynamics structure as for integrated models in Category 6 and statistical catch-at-age models. They may allow for multiple areas and multiple growth patterns, environmental covariates for various processes, internal estimation of natural mortality and growth (e.g. possibly by using age–length keys and length distribution data as inputs, rather than merging these earlier to input as catch-at-age data), and time-varying processes. With high age data quality, they can approach a VPA configuration. In weak data situations, use of fixed parameters or priors mimics a simple age-structured production model.*

b. *Min data – catch and an abundance index (some missing data allowed). Some implementations allow the catch data to be separated into landings and discards.*
c. *Typical data – catch, multiple abundance indices, age and/or length data. May also include tag-recapture data to assist estimate F, M and its age dependence and movement, and also stock structure (including genetics) data to estimate proportions of different stocks present.*
d. *Examples – Stock Synthesis, CASAL, IWC minke whale multistock models*

Note that some of the features used in differentiating these model categories will occur at multiple levels. For example, multi-area configurations with stock-structure data might be used in a surplus production model. In other cases, fully integrated analysis models can be configured to operate with limited data and be configured to perform as biomass dynamics models.

Discussion:

Categorization Scheme:

- It would help to add a part "e." under each of the model categories, with the aim of crossing assessment categories with possible management advice under typical data for that model category (in particular, reference points that can be calculated for management)
- Multispecies models were considered to be missing from the symposium topics. The presenters explained that multispecies models are not a main focus of SISAM, but that a session on such models would be included in the 2013 symposium.
- Underlying model assumptions should be stated in a little more detail under part "a. Structure"

Selection of datasets

In order to facilitate comparison studies at the 2013 symposium workshops, SISAM plans to assemble 10 to 12 datasets during 2012. One of the SISAM ToRs proposes that WGMG in 2012 helps with this task. SISAM considered it useful already to start a discussion with WGMG regarding these datasets this year, in order to gather the views of the scientists present at this year WGMG meeting. The datasets should be representative of the whole range of situations covered by the SISAM initiative. It was also felt that no more than 10 to 12 datasets should be considered, as the aim is to apply different models and techniques to the same datasets in order to try to understand the properties and consequences of applying different methodologies, and to get a feeling for what might be considered as best practice. The following range of features was deemed to be relevant in the selection of datasets:

- Data rich vs. poor
- Landings vs. discards uncertainty
- Short vs. long lived species
- Low vs. high recruitment variability (e.g. sudden recruitment outbursts as can be seen in horse mackerel)
- Stocks that have dropped to very low abundance and then recovered (e.g. Norwegian spring-spawning herring)
- Stocks with multiple components (e.g. North Sea herring)
- Stocks currently under ICES WGNEW

It was also considered very important for the success of the work to be conducted with the selected datasets that scientists knowledgeable about those stocks and data, and interested in collaborating in this work, be identified.

10 Conclusions and Recommendations

ToR 1a

Data screening techniques prior to the selection of stock assessment models

Background

Useful for:

- Exploring and demonstrating data features.
- Checking for consistency within and between data sources.
- Providing ball-park trends to be expected from assessment model.
- Understanding behaviour of assessment models.

Not used enough in assessment reports.

Recommendations

It is recommended that showing outputs from data pre-screening techniques that proved informative should be a standard requirement in ICES stock assessment reports.

ToR 1b

Diagnostics to evaluate model fit (including measures of retrospective bias), and how these can be used to help refine models where appropriate

Background

Retrospective indices:

- Potential for developing threshold levels in retrospective indices beyond which inaccuracy would be unacceptably large.
- Needs further work:
  - Checking behaviour of retrospective index under alternative assumptions for generating simulated data.
  - Testing under a variety of simulated population vs. assessment model combinations.

Incorporating estimates of sampling variability in assessments:

- Potential for using estimates of survey sampling variability as inputs to XSA to weight individual survey data points (by year and age).
- Needs further work to check statistical assumptions.

Recommendations

It is recommended that estimates of survey sampling variance always be calculated. Where appropriate, the inverse of survey estimates of sampling variance should be incorporated as a maximum weighting for corresponding survey data points.

ToR 1c

Guidance for deciding how complex a stock assessment model needs to be (e.g. how much to process/aggregate inputs; utility for advice)

Simple to more Complex

Background

Pros:

- Fuller use of available data/more biological realism (anglerfish example).
- Simpler models give deceptively small confidence intervals.
- Allows more flexibility, for example:
    - Can investigate impact of changing the selection pattern (megrim example).
    - More appropriate modelling of landings and discards (North Sea cod example).

Cons:

- Danger of over-parameterization.

There is an overarching concern that "acceptable" model choice approaches are followed and model-fitting diagnostics are obtained (e.g. residuals are broadly random).

- The Ecoknows perspective is that model specification should be driven by realistic biological and population dynamics assumptions, and not data availability alone.
- Although residual patterns may not be corrected for (e.g. autocorrelation), it is important to be aware of them, particularly in the context of MSE, to ensure that pseudo-data have the same properties as actual historic data.

Recommendations

- Consider using AIC in a frequentist or DIC in a Bayesian setting, for example, to guard against over-parameterization. Take care however to consider whether the data concerned are independent, as these approaches assume.
- When introducing random effects terms, the statistical properties assumed should be checked to the extent possible, e.g. when appropriate through a runs test to check for randomness.

Harvest control rules that use fewer data

Background

- Harvest control rules which use fewer data (e.g. only survey indices of abundance) have been found for certain stocks that outperform what actually happened in the past in terms of actual removals (based on a complex assessment) in almost every respect, particularly interannual variability in catch and fishing mortality.
- The MP testing framework could be used to evaluate the loss (in terms of more conservative catch limits) of reducing the amount of data collected.

<u>*Recommendation*</u>

- It is recommended that the approach used to evaluate simple management procedures, described in Annex 5, WD 7 be developed further as a possible framework for investigating the value of information.

ToR 1d

Implications and treatment of correlated errors

Background

The investigations focused on year effect in surveys and estimating stock recruit relationships taking autocorrelation in recruitment into account by considering AR processes for residuals. The main conclusions were:

- Important to account for correlated errors to better reflect the information content of data.
- Better modelling of survey data before asking assessment model to "figure it out".
- Trying to estimate the 1^{st} and 2^{nd} order parameters of an auto-regressive process can lead to strange behaviour, requiring the imposition of a penalty to ensure residuals sum to zero.
- There is a big difference in SR models estimated using the AR(1) or AR(2) formulations.
- There is potential that an autocorrelated recruitment error structure can confound the stock–recruit signal.
- Work in progress, so no firm recommendation can be made at this stage.

ToR 1e

Integration of uncertainty (including accounting for retrospective patterns) in advice

No work presented explicitly addressing this topic (although there are links to work presented elsewhere in this report – e.g. see Section 6).

Review approaches for standardizing commercial cpue (available techniques and pitfalls)

Background

- A review is provided.
- Example GLM application based on the Tweedie distribution given.

ToR 3

Provide guidelines for calculating MSY reference points in a varying and stochastic environment

Background

Other ICES WGs have dealt with this (WKFRAME, WKFRAME2, SGMAS). The study presented was limited to use of SURBA+ as the assessment model, and to 3PS cod and American plaice.

- Measurement error bias in fish stock spawner-recruitment models:

    - Simulation analyses showed that measurement error in SSB, if substantial, could have a large impact on MSY reference points, and parameters such as S50% (SSB value at half asymptotic recruitment), calculated from the estimated stock–recruit relationship.
    - Could provide guidelines for use of more robust reference points.
    - Needs further work.
- Maximum sustainable yield when recruitment productivity varies:
    - Bousquet *et al.* (2008) concluded that their study "reinforced the conviction shared by numerous researchers that biological reference points calculated in a deterministic framework can be far from optimal in stochastic settings".
    - The study presented during the meeting found that:
        - The amount of process error had little effect on mean MSY reference points, which differs from the conclusions of Bousquet *et al.* (2008); however upper and lower percentiles were affected.
        - Constraining multiplicative process errors to have a geometric mean of one (i.e. average of the log-process errors is zero) makes a difference.
        - Needs further work.

Work in progress, so no firm recommendation can be made at this stage.

11 Future of WGMG and ToRs for 2012

Background

Discussion on the future direction of WGMG came up during the 2009 WGMG meeting (Section 10.2 in ICES, 2009). This discussion arose from a change in approach to setting ToRs that was tried for the 2009 WGMG, where the ToRs were developed intersessionally by the WGMG chair in consultation with benchmark and assessment Working Group chairs. Although the intention of the new approach was reasonable (ensuring the relevance of WGMG output to the direct requirements of forthcoming benchmark and assessment meetings), it was felt by WGMG at the time to have failed for several reasons:

- there was considerably less buy-in from benchmark and assessment chairs than had been anticipated, with three out of the five individuals submitting requests being themselves WGMG members;
- the consultation process meant that ToRs were finalized only during summer immediately preceding the WGMG meeting in autumn, allowing a very short period for decisions to be made about attendance – this is problematic for institutions that plan their travel budgets at the start of the year, and it was clear from this experience (only 6 members attended, 1 part-time) that a WGMG meeting without defined ToRs received a low priority for many who may have otherwise attended.

The role of WGMG in 2010 was largely in support of preparations for WKADSAM, a workshop on reviews of recent advances in stock assessment models worldwide (ICES, 2010b) – this workshop effectively launched the ICES Strategic Initiative on Stock Assessment Methods (SISAM). The support given by WGMG to the workshop was by correspondence, so WGMG did not physically meet as a working group during 2010. However, during the WKADSAM meeting, the members of WGMG present met informally to agree a set of ToRs for 2011, based on supporting the ICES SISAM, and on topics that needed further work, derived from an analysis of benchmark reports presented to this WKADSAM meeting. The venue (Vigo) and length of meeting (10 days) was also agreed. In order to deal with the concern about ToRs being developed and released well in advance, these ToRs were agreed by SCICOM and released before the start of 2011. It was hoped that developing the ToRs in this manner would keep the interest of WGMG members and maintain relevance to ICES needs.

In the event, only a small number of experienced members indicated they would attend the 2011 meeting after the release of the ToRs, and it was felt that it would be useful to reopen the discussion on the future direction of WGMG. An attempt was made, through e-mail circulation, to get feedback on the topic from members not involved with or attending the 2011 meeting in order to contribute ideas to the discussion to be held during the meeting, but only a single response was received prior to the meeting. Nevertheless, some time was spent during the 2011 meeting on this topic, and a summary of the discussion follows.

Discussions during the meeting

In order to attract a greater number of people to attend and contribute to the ToRs of the meeting, it was felt that a number of factors needed to be considered, including (not in order of priority):

- a re-assessment of the length of the meeting (many considered 10 days too long)
- finding ways of improving attendance and ensuring participation of a critical mass of experts
- better advertising of the existence of the group and its activities
- enhancing the potential for publication
- enhancing the potential for networking and collaboration
- enhancing student participation to assist develop their expertise in the field
- ToRs being made available well in advance of the meeting (at least at the beginning of the year)
- the format of future meetings should address benefits to ICES as well as benefits to participants.

In order to address benefits to ICES, it was felt that a stronger direction from benchmark and assessment WG chairs was needed in developing ToRs, but that these ToRs should not be addressing short-term quick-fixes to assessment problems (more appropriate to benchmark meetings), but rather problems of methodology that may need a longer period of time to address, and quite possibly also be common across a number of WGs. This avoids the problem of ToRs being made available only shortly before a meeting. It was also felt that, along with developing ToRs, experts or particular groups of people should also be identified that would be able to work on these ToRs – in this way, the probability of WGMG being able to address its ToRs would be greatly enhanced. Both the identification of ToRs and the selection of groups to address each of these could take place at the ICES WG Chairs meeting at the start of the year.

In order to address benefits to participants, several ideas were put forward.

- Fewer, more focused ToRs would encourage the participation of several people/groups of people working on the same ToR, but addressing it from different angles, fostering both collaboration and a cross-pollination of ideas. A particular problem during the meeting in 2011 was that, on the whole, the ToRs were addressed from a single viewpoint. In contrast one would want to see a greater fraction of the WG time spent in separate meetings with such smaller groups using the opportunity of meeting together to advance on their problem.
- A "PhD day" that forms part of the meeting, where several students present work that could help address the ToR, would liven up debate and benefit both the students and experts. A pre-circulation of abstracts would ensure that a close link to the ToRs is maintained, and their availability through a SharePoint site (along with contributions from other participants) would encourage collaboration.

Methodological issues are not unique to ICES, and it was felt that ICES may benefit from collaboration with other groups within ICES and with other organizations. For example, it was suggested that:

- WGMG could team up with WGSAM (the multispecies WG) as a means of pooling expertise through either back-to-back meetings (but this would not help with the length of the meeting for some people), or through meetings held in parallel, but with some shared sessions. The idea of being closer to WGSAM was favoured by some WG members.

- WGMG could team up with methods groups from other organizations (e.g. ICCAT has shown interest) and hold a joint meeting every second year dealing with shared problems, with interim years dealing only with ICES issues.
- WGMG could attach itself to a major meeting every four years (such as the World Fisheries Congress) as a means to involve experts attending these meetings. For example, in addition to the MSE session at the WFC meeting in Edinburgh in 2012 being an opportunity to raise this idea, this discussion could also look at linking other assessment methodology sessions held more regularly.

Proposed way forward for 2012

It may take some time to develop ToRs aligned to ICES needs and to identify groups to address each of these (e.g. through the ICES WG Chairs meeting at the start of 2012), resulting in these ToRs being approved only well into 2012 and, hence, being too late for the 2012 WGMG meeting. It is therefore proposed that for the WGMG meeting in 2012, WGMG is more closely aligned to the ICES SISAM initiative. This would buy more time for the ToR for 2013 to be developed in a way that would better service the needs of ICES assessment Working Groups and Benchmark meetings (on methodological issues), allow groups of scientists to be targeted to work on these ToRs, and be planned in such a way as to attract the range of experts that could advance the field. Aligning WGMG with SISAM for 2012 would also help maintain the relevance of WGMG to ICES, and could act as a focal point for a strong ICES contribution to the 2013 symposium.

The following ToRs are therefore proposed for 2012:

1) Assemble 10–12 datasets from ICES that characterize the breadth of life history strategy, data quality, population dynamics, and assessment problems.
2) Prepare a publication (to be presented to the SISAM symposium), using these datasets, that explores providing guidelines on simulation testing of assessment models.
3) In preparation for the SISAM symposium and building on WKADSAM, pre-test/challenge a selection of stock assessment models on the assembled datasets.
4) Using these tests, and the newly developed model categorization scheme, highlight the weaknesses and strengths of the ICES approach and the current portfolio of stock assessment models used by ICES.

Proposed venue: IPIMAR, Lisbon, Portugal.

Proposed dates: 5 days in the period 24 September – 12 October 2012 (e.g. 1–5 October).

12 References

Aarts, G., and Poos, J. J. 2009. Comprehensive discard reconstruction and abundance estimation using flexible selectivity functions. ICES Journal of Marine Science, 66: 763–771.

ADMB Project. 2009 AD Model Builder: automatic differentiation model builder. Developed by David Fournier and freely available from admb-project.org.

Bousquet, N., Duchesne, T., and Rivest, L.-P. 2008. Redefining the maximum sustainable yield for the Schaefer population model including multiplicative environmental noise. J. Theor. Biol., 254: 65–75.

Brodziak, J., and Legault, C. 2005. Model averaging to estimate rebuilding targets for overfished stocks. Canadian Journal of Fisheries and Aquatic Sciences, 62: 544–562.

Casey, J. 1996. Estimating discards using selectivity data: the effects of including discard data in assessments of the demersal fisheries in the Irish Sea. Journal of Northwest Atlantic Fishery Science, 19: 91–102.

Chen, Y., Xu, L., Chen, X., and Dai, X. 2007. A simulation study of impacts of at-sea discarding and bycatch on the estimation of biological reference points F0.1 and Fmax. Fisheries Research, 85: 14–22.

Dickey-Collas, M., Pastoors, M. A., and van Keeken, O. A. 2007. Precisely wrong or vaguely right: simulations of noisy discard data and trends in fishing effort being included in the stock assessment of North Sea plaice. ICES Journal of Marine Science, 64: 1641–1649.

Fernández, C., Cerviño, S., Pérez, N., and Jardim, E. 2010. Stock assessment and projections incorporating discard estimates in some years: an application to the hake stock in ICES Divisions VIIIc and IXa. – ICES Journal of Marine Science, 67: 1185–1197.

ICES. 2008a. Report of the Working Group on Methods of Fish Stock Assessment (WGMG). ICES Document CM 2008/RMC:03. 147 pp.

ICES. 2008b. Report of the Study Group on Management Strategies (SGMAS), 17-21 November 2008, Lisbon, Portugal. ICES CM 2008/ACOM:24. 74 pp.

ICES. 2009. Report of the Working Group on Methods of Fish Stock Assessment (WGMG), 20–29 October 2009, Nantes, France. ICES CM 2009/RMC:12. 85 pp.

ICES. 2010a. Report of the Workshop on Implementing the ICES Fmsy framework , 22-26 March 2010, Copenhagen, Denmark. ICES CM 2010/ACOM:54. 83 pp.

ICES. 2010b. 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.

ICES. 2011. Report of the Workshop on Implementing the ICES Fmsy Framework (WKFRAME-2), 10–14 January 2011, ICES, Denmark. ICES CM 2011/ACOM:33. 110 pp.

Jardim, E., Cerviño, S., and Azevedo, M. 2010. Evaluating management strategies to implement the recovery plan for Iberian hake (*Merluccius merluccius*); the impact of censored catch information. ICES Journal of Marine Science, 67: 258–269.

Maunder, M. M., Punt, A. E., 2004. Standardizing catch and effort data: a review of recent approaches. Fisheries Research, 74: 141–159.

Method, R. D. 2005. Technical description of the stock synthesis II assessment program: Version 1.17, 54p.

Myers, R. A., and Cadigan, N. G. 1995. Statistical analysis of catch-at-age data with correlated errors. Can. J. Fish. Aquat. Sci., 52 (6): 1265–1273.

Punt, A. E., Smith, D. C., Tuck, G. N., and Methot, R. D. 2006. Including discard data in fisheries stock assessments: two case studies from south-eastern Australia. Fisheries Research, 79: 239–250.

Shono, H. 2008a. Application of the Tweedie distribution to zero-catch data in CPUE analysis. Fisheries Research, 93: (1–2), 154–162.

Shono, H. 2008b. Confidence interval estimation of CPUE year trend in delta-type two-step model. Fisheries Science, 74: (4), 712–717.

Sissenwine, M. P., Shepherd, J. G. 1987. An Alternative Perspective on Recruitment Overfishing and Biological Reference Points. Can. J. Fish. Aquat. Sci., 44: 913–918.

Tweedie, M. C. K. 1984. An index which distinguishes between some important exponential families. *In:* Statistics: Applications and New Directions. Ed. By Ghosh, J. K., and Roy, J. Proceedings of the Indian Statistical Institute Golden Jubilee International Conference. Indian Statistical Institute, Calcutta, pp. 579–604.

Venables, W. N., Dichmont, C. M., 2004. GLMs, GAMs and GLMMs: an overview of theory for applications in fisheries research. Fish. Res., 70 (2004): 319–227.

Williams, E. H. 2002. The effects of unaccounted discards and misspecified natural mortality on harvest policies based on estimates of spawners per recruit. North American Journal of Fisheries Management, 22: 311–325.

Annex 1: List of participants

Name	Address	Telephone/Telefax	E-mail
Esther Abad	Instituto Español de Oceanografía Centro Oceanográfico de Vigo Cabo Estai - Canido PO Box 1552 36200 Vigo (Pontevedra) Spain	+34 986 492 111	esther.abad@vi.ieo.es
José De Oliveira **Chair**	Centre for Environment, Fisheries and Aquaculture Science (Cefas) Lowestoft Laboratory Pakefield Road NR33 0HT Lowestoft Suffolk UK	+44 1502 527727 +44 1502 524 511	jose.deoliveira@cefas.co.uk
Timothy Earl	Centre for Environment, Fisheries and Aquaculture Science (Cefas) Pakefield Road NR33 0HT Lowestoft Suffolk UK	+44 (0) 1502 521303	timothy.earl@cefas.co.uk
Carmen Fernández	Instituto Español de Oceanografía Centro Oceanográfico de Vigo Cabo Estai - Canido PO Box 1552 36200 Vigo (Pontevedra) Spain	+34 986 492111 +34 986 498626	carmen.fernandez@vi.ieo.es
Pavel Gasyukov	AtlantNIRO 5 Dmitry Donskogo Street RU-236000 Kaliningrad Russian Federation	+7 4012 225 257 +7 4012 219 997	pg@atlant.baltnet.ru
Monica Mandado	Instituto de Investigaciones Marinas - CSIC Eduardo Cabello 6 ES-36208 Vigo (Pontevedra) Spain	+34 986 231930	mandado@iim.csic.es
Andrey Mikhailov	Russian Federal Research Institute of Fisheries & Oceanography (VNIRO) 17 Verkhne Krasnoselskaya 107140 Moscow Russian Federation	+7 849926459091 +7 84992649078	mikhailov1984@gmail.com
Lionel Pawlowski	Ifremer Lorient Station 8, rue François Toullec 56100 Lorient France	+33 2 97 87 38 46 +33 2 97 87 38 36	lionel.pawlowski@ifremer.fr

Name	Address	Telephone/Telefax	E-mail
Paz Sampedro	Instituto Español de Oceanografía Centro Oceanográfico de A Coruña Paseo Marítimo Francisco Vázquez, 1 15001 A Coruña Spain	+34 981 218253	paz.sampedro@co.ieo.es
Antonio Vázquez	Instituto de Investigaciones Marinas - CSIC Eduardo Cabello 6 E-36208 Vigo Spain	+34 986 231930	avazquez@iim.csic.es
Attending Working Group Meeting as Part-time Participant			
David Miller	Wageningen IMARES PO Box 68 1970 AB IJmuiden Netherlands		david.miller@wur.nl
Doug Butterworth	University of Cape Town Dept of Mathematics & Applied Mathematics 7701 Rondebosch South Africa	21 650 2343	doug.butterworth@uct.ac.za
Helena F. Geromont	University of Cape Town Dept of Mathematics & Applied Mathematics 7701 Rondebosch South Africa	+27 21 650 3191 +27 21 650 2334	Helena.Geromont@uct.ac.za
Samuel Subbey	Institute of Marine Research Nordnes PO Box 1870 5817 Bergen Norway	+47 5523 5383 +47 5523 8687	samuel.subbey@imr.no
Yuri A. Kovalev	Knipovich Polar Research Institute of Marine Fisheries and Oceanography(PINRO) 6 Knipovitch Street 183763 Murmansk Russian Federation	+7 8152 472 469 +7 8152 473 331	kovalev@pinro.ru
Working by correspondence attending by WebEx/Skype			
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
Anders Nielsen	DTU-Aqua Jægersborg Allé 1 2920 Charlottenlund Denmark	+45 35 88 33 00	an@aqua.dtu.dk

Annex 2: Agenda

Monday 10th October:

10:00 Welcome and introductions
Brief introduction to Southern horse mackerel data (Gersom Costas)
Presentation 1: ToR 1a – Timothy Earl (rap: Andrey Mihailov)

11:00 Coffee

11:00–13:00: No plenary

13:00–14:00: Lunch

14:00 Presentation 2: ToR 1b – Antonio Vázquez (rap: Monica Mandado)
Presentation 3: ToR 1b –Pavel Gasyukov (rap: Timothy Earl)

15:30–17:30: No plenary

15:45 Coffee

Tuesday 11th October:

09:00 Presentation 4: ToR 1c – Paz Sampedro (rap: Lionel Pawlowski)
Presentation 5: ToR 1c – Carmen Fernández (rap: Yuri Kovalev)

10:30–13:00: No plenary

10:45 Coffee

13:00–14:00: Lunch

14:00 Presentation 9: ToR 2 – Lionel Pawlowski (rap: Paz Sampedro)

15:00–17:30: No plenary

15:45: Coffee

Wednesday 12th October:

09:00 Presentation 7: ToR 1e (ECOKNOWS) – Samu Mäntyniemi (rap: Carmen Fernández)
Other possible presentations under ToR 1e (to be confirmed)

10:30–13:00: No plenary

10:45 Coffee

13:00–14:00: Lunch

14:00 Presentation 6a&b: ToR 1d, 3 – Noel Cadigan (rap: Samuel Subbey)
Presentation 8: ToR 2 – Pavel Gasyukov (rap: Antonio Vázquez)

15:30–17:30: No plenary

15:45 Coffee

Thursday 13th October:

09:00 Presentation 10: ToR 1c – Anders Nielsen (rap: Helena Geromont)

10:00–13:00: No plenary

10:45 Coffee

13:00–14:00: Lunch

14:00–16:00: No plenary

15:45 Coffee

16:00 Presentation 12: ToR 4 – Doug Butterworth/Carmen Fernández (rap: José De Oliveira)
Presentation 11: ToR 1c – Helena Geromont/Doug Butterworth (rap: David Miller)

Evening: Group Dinner

Friday 14th October:

09:00 Finish off discussions on Presentation 11
Finish off discussions on Presentation 12
ToR 1e – discussion led by David Miller

10:45 Coffee

13:00–14:00: Lunch

14:30 Presentation by Noel Cadigan (rap: Timothy Earl)
Finish off discussion on Presentation 8

15:45 Coffee

Saturday 15th October:

09:00 Discussion of Recommendations so far
Discussion of future direction of WGMG

10:45 Coffee

12:00 No plenary for remainder of day.

Sunday 16th October:

Free Day (optional outing to Cangas and walk, or Cies)

Monday 17th – Wednesday 19th October:

Consists of: follow-up presentations, work on report, and discussion of recommendations and suggested ToR for 2012.

Annex 3: WGMG terms of reference for the next meeting

The **Working Group on Methods of Fish Stock Assessments** (WGMG) chaired by José De Oliveira, UK, will meet in Lisbon, Portugal, 1–5 October 2012 to:

a) Assemble 10–12 datasets from ICES that characterize the breadth of life history strategy, data quality, population dynamics, and assessment problems.

b) Prepare a publication (to be presented to the SISAM symposium), using these datasets, that explores providing guidelines on simulation testing of assessment models.

c) In preparation for the SISAM symposium and building on WKADSAM, pre-test/challenge a selection of stock assessment models on the assembled datasets.

d) Using these tests, and the newly developed model categorization scheme, highlight the strengths and weaknesses of the ICES approach and the current portfolio of stock assessment models used by ICES.

WGMG will report by 1 December 2012 (via SSGSUE) for the attention of the SCICOM.

Supporting information

Priority	The work of this group is essential to ICES to progress in the development of methods for fish stock assessment and advice.
Scientific justification	The overarching plan of WGMG is to improve service to the needs of ICES assessment Working Groups and Benchmark meetings (on methodological issues), identifying groups of scientists to work on the ToRs and to plan it in such a way as to attract the range of experts that could advance the field. It may take some time to develop ToRs aligned to ICES needs and to identify groups of scientists to address each of these (e.g. through the ICES WG Chairs meeting at the start of 2012), resulting in these ToRs being approved only well into 2012 and, hence, being too late for the 2012 WGMG meeting. It is therefore proposed that the focus of the WGMG meeting in 2012 is more closely aligned to the ICES SISAM initiative. This would buy more time for the ToRs for 2013 to be developed along the lines mentioned at the beginning of this paragraph. The ICES SISAM initiative and associated symposium planned for 2013 are important drivers for advancing the incorporation of relevant develpoments in stock assessment methods into the ICES advisory system so as to ensure ICES scientists can apply the best methods when developing management advice, and can make better use of available resources. Aligning WGMG with SISAM for 2012 would help enhance the relevance of WGMG to ICES, and could act as a focal point for a strong ICES contribution to the 2013 symposium. WGMG help with the selection of 10–12 datasets (ToR 1), representative of a wide range of situations, is considered a key contribution to the SISAM initiative in order to facilitate comparison studies during the 2013 SISAM symposium workshops. The aim is to apply different models and techniques to the same dats sets in order to try to understand the properties and consequences of applying different methodologies and to get a sense for what might be considered as best practice. These datasets should cover a range of features: data rich vs. data poor, landings vs. discard uncertainty, short- vs. long-lived species, low vs. high recruitment

	variability, etc. Another important aspect of the SISAM initiative is the simulation testing of assessment models, and ToR 2 aims to develop guidelines, through the preparation of a publication, on how such simulation analyses should best be conducted. This could be done based on generated datasets. However, the approach here would not be to repeat the blind-testing of assessment models performed in the past (e.g. NRC study 1998, WGMG 2004); rather the approach would be to condition simulation studies on the actual data for the stocks under consideration. ToR 3 looks both to coordinating a strong ICES contribution to the SISAM symposium by pre-testing a selection of models, and to use the experience to provide feedback to the SISAM steering group that could be used to improve planning for the 2013 symposium workshops. Based on these tests and using the model categorization scheme, ToR 4 aims to highlight the strengths and weaknesses of approaches and models currently used by ICES.
Resource requirements	None.
Participants	Research scientists involved in stock assessment methods from the ICES area and elsewhere in the world.
Secretariat facilities	None, other than formatting and publishing of the final report.
Financial	None.
Linkages to advisory committees	ACOM has strongly supported the work of this group. WGMG will report to ACOM in 2012.
Linkages to other committees or groups	WGMG will report to SCICOM in 2012. WGMG involved with the ICES Strategic Initiative on Stock Assessment Methods (SISAM).
Linkages to other organizations	NAFO, ICCAT.

Annex 4: Recommendations

Recommendation	Adressed to
1. It is recommended that showing outputs from data pre-screening techniques that proved informative should be a standard requirement in ICES stock assessment reports	ICES Assessment WGs
2. It is recommended that estimates of survey sampling variance always be calculated. Where appropriate, the inverse of survey estimates of sampling variance should be incorporated as a maximum weighting for corresponding survey data points.	WGISDAA, WGIPS, WKTSBLUES, WGISUR, SGNEPS, WGBIFS, IBTSWG, WKMSPA, WGMEGS, WGBEAM, WGNEACS, ICES Assessment WGs
3. Consider using AIC in a frequentist or DIC in a Bayesian setting, for example, to guard against over-parameterization. Take care however to consider whether the data concerned are independent, as these approaches assume.	ICES Assessment WGs
4. When introducing random effects terms, the statistical properties assumed should be checked to the extent possible, e.g. when appropriate through a runs test to check for randomness.	ICES Assessment WGs
5. It is recommended that the approach used to evaluate simple management procedures, described in Annex 5, WD 7 be developed further as a possible framework for investigating the value of information.	ACOM

Annex 5: Working Documents

Working Document 1

Data Screening Techniques Prior to the Selection of Stock Assessment Models

T. Earl, J. De Oliveira and C. Darby

To analyse the variety of relevant plots currently in use, a number of the most recent reports from ICES working groups[1] were checked for data screening plots and a selection of these are shown in this document. To include plots from institutions outside ICES, the most recent NOAA GARM report, WCPCF skipjack tuna and ICCAT reports were also examined.

The emphasis of this search was primarily concerned with catch-at-age and length, and survey at age data, but some illustrations of spatial data are also shown in the final section.

The majority of data screening plots are motivated by addressing a particular question or concern about the stock, whereas the ToR is to investigate plots that may be illuminating for a variety of stocks. Some general points that it may be important to display in a plot across a wide range of stocks are:

1) Is the level of catch consistent with expectations based on previous years or management plans?

2) What is the age structure of the catch; does this change over time?

3) What do the indices indicate about changes in the stock level and structure over time?

4) Is each index showing consistent trends across the age-range?

5) If there is more than one index, are the indices showing consistent trends?

6) Do the survey areas reflect the present location of the stock and fishing effort appropriately?

7) What effect would the new data be expected to have on the assessment?

Catch

The simplest displays of catch are time-series of landings/catches by weight or numbers such as Figure 1, showing catch numbers by year, with the mean catch and selected percentiles added. Variations on this plot include using stacked bars to indicate catches from different regions, different gears or different landing countries, such as Figure 2. This plot also shows how the catches compare to the TAC for each year.

Typically, catch data are available as a catch-at-age matrix, and Figures 3, 4 and 5 illustrate ways of displaying this. Figure 3 shows bar charts of the age distribution each year, with a common x- and y-scale across years to aid comparison. Figure 4 shows the catch-at-age data as a bubble plot. This has the advantage compared to the previous plot of allowing individual cohorts to be clearly seen (as diagonal lines), but

[1] AFWG, HAWG, WGEAWESS, WGIAB, WGWIDE, WKBALTEEL, WKBENCH, WKCOD, WKDEEP, WKFLABA, WKFLAT, WKWIDE

this type of plot needs to make clear whether catch is related to area or diameter of the bubble. Figure 5 shows a variation on Figure 4, with the bubbles indicating proportion of the catch in that year at each age, while the bars along the top indicate the size of the total catch, broken down by landing country. An advantage of separating the data in this way is that it reflects the different steps in compiling catch-at-age data, i.e. landings weight is based on all fish landed, whereas the age distribution is only based on a sample of landings.

Figure 6 shows a series of boxplots showing the catches at different ages, compiled from a number of years. This gives the impression of the age structure of the catch without showing any temporal trends. The averaging over a number of years may help remove cohort effects, but may also be hiding information about gear selectivity changing.

Figure 7 shows the proportions of catch by gear over time, which may be more useful than absolute catches in cases where the magnitude of catch changes so much that it would be hard to see the composition of the catch in years where the total catch is small.

In general, catch at length data are used less frequently in ICES assessments, but where this is used bar charts of length classes can be plotted in the same way as age classes. Bubble plots are less applicable to this type of data, because cohorts are not clearly displayed.

The only single figure that displays both the age structure and total landings is Figure 5, which would make it a good candidate for including in catch data screening. It could be further enhanced by adding the TAC set for each year in a similar manner to Figure 2 to indicate the relationship between TAC and landings.

Catch at length

Many ICES assessments are based on catch-at-age data, but Figure 8 shows an example of cpue at length data from outside ICES. Using an age–length relationship, an individual cohort has been highlighted in red, and other cohorts have been added in grey. It isn’t clear from the plots whether one should refer to the point where the red line crosses the x-axis on each plot, or to the point where it touches the histogram, or some other point. The use of the same scale in each subplot makes it easy to spot the decrease in cpue in the last year, but makes it harder to see the age structure in the years with largest and smallest numbers.

Figure 9 combines a length and age distribution for the current year, although the legend and scale on the length distribution are unclear, the overall distribution is clear

Catch data Frequency

Figure 10 gives an indication of the amount of length data from a variety of sources that are used in an assessment. A similar plot for age based data could show number of individual aged. The maximum for each fleet is shown on the right hand side (possibly this could also be a horizontal bar chart), and within each fleet the number of samples taken relative to the maximum year for that fleet is shown as a bar chart. This gives a clear visualization of the source of length measurements, but may not be useful unless compared to the relative catch contribution of each fleet.

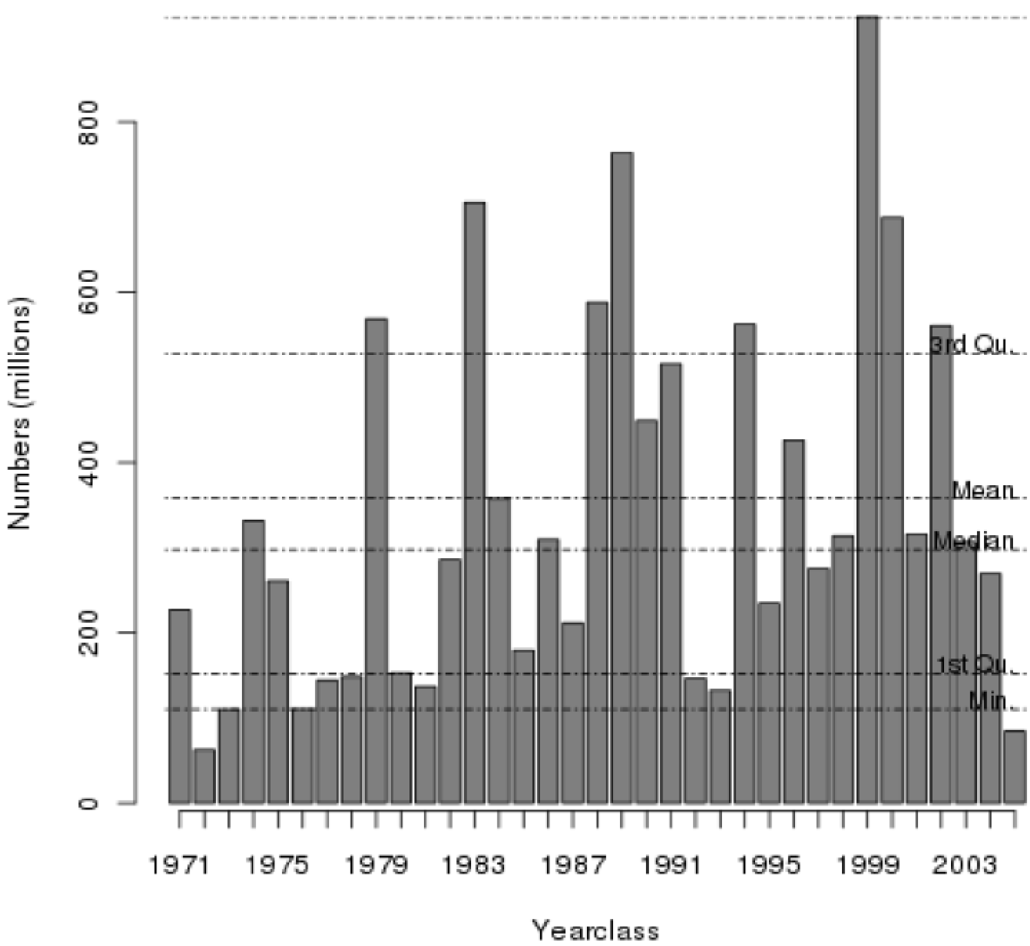


Figure 1. Catch numbers per year (WKBENCH 2011). Summary statistics (0, 25, 50, 75, 100%iles and mean) refer to 1973–2004, i.e. disregarding the initial 4 years, and last year.

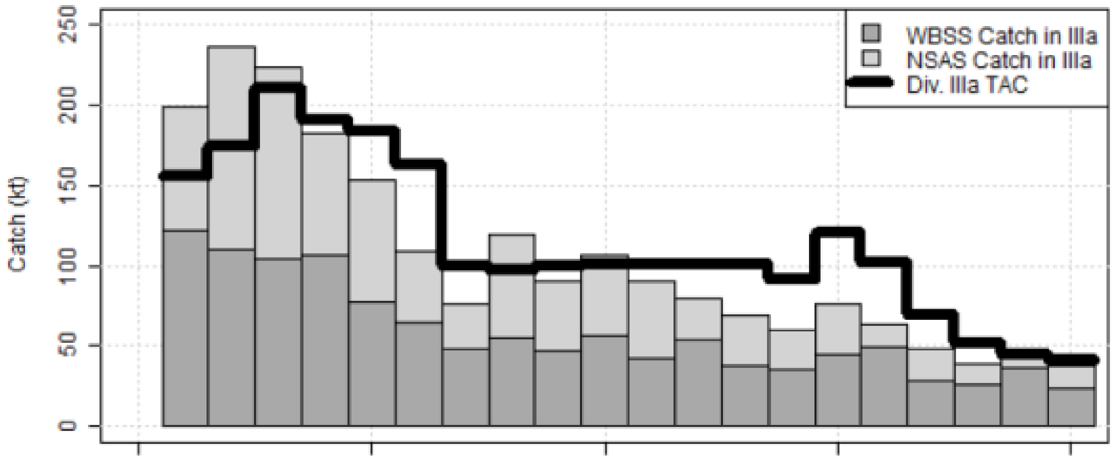


Figure 2. Catch tonnage, including TAC (from HAWG 2011).

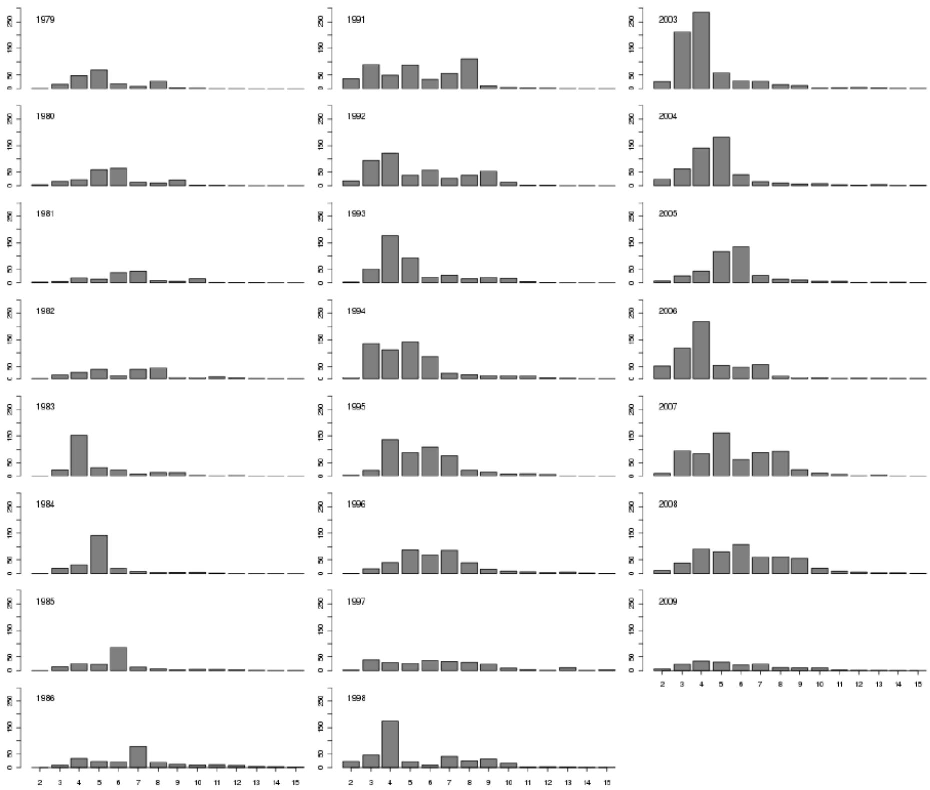


Figure 3. Catch numbers by age for a selection of years (from WKBENCH 2011).

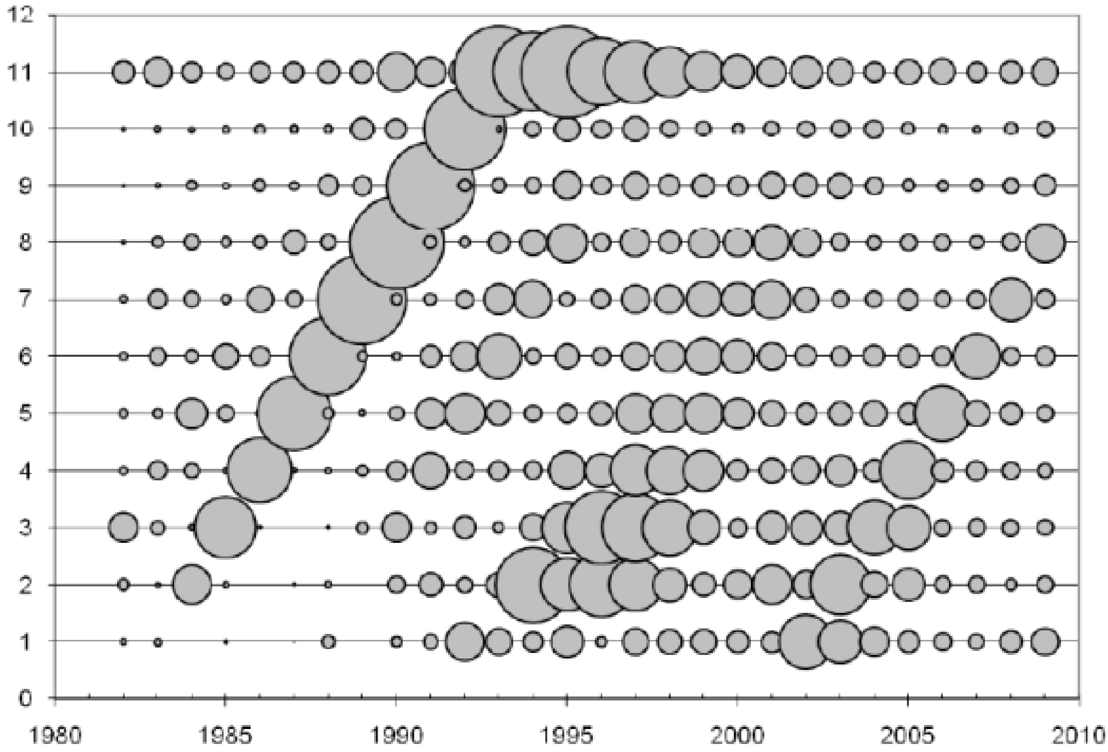


Figure 4. Catch-at-age data with area proportional to catch size. Note age 11 is a plus group (WGWIDE 2010).

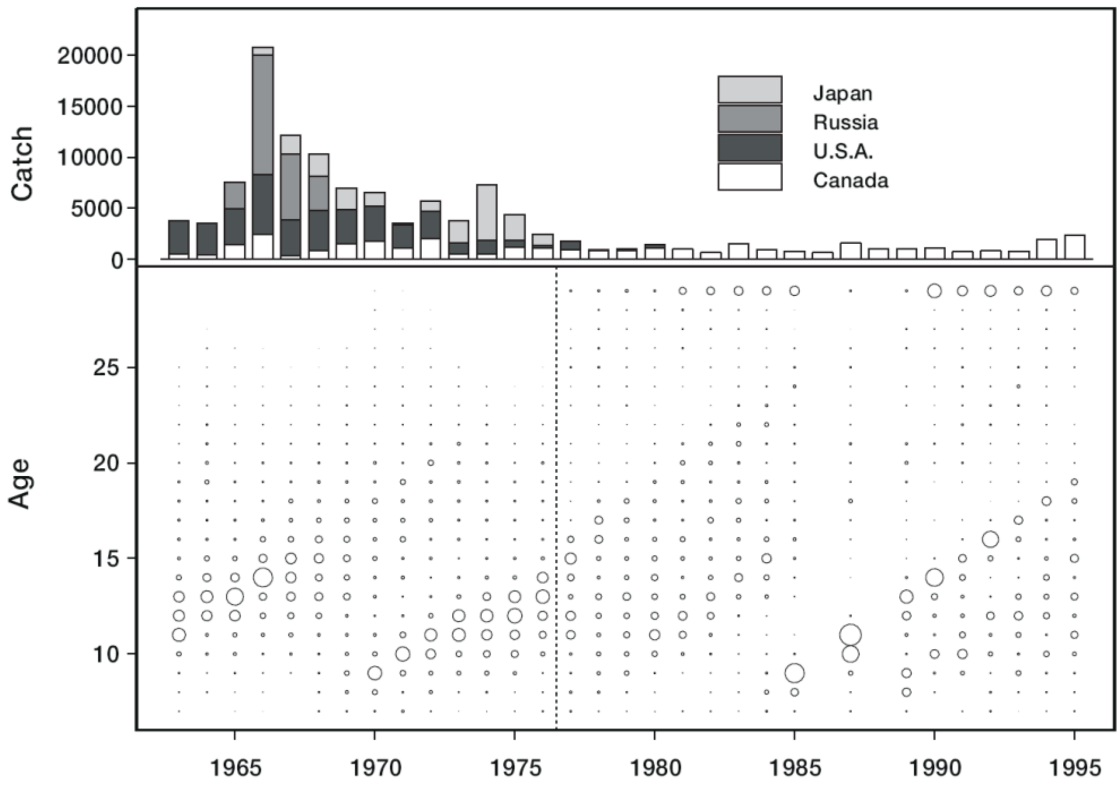


Figure 5. Catch data by landing country, and relative proportions at age. Dotted line indicates change in aging method. From Richards (2007).

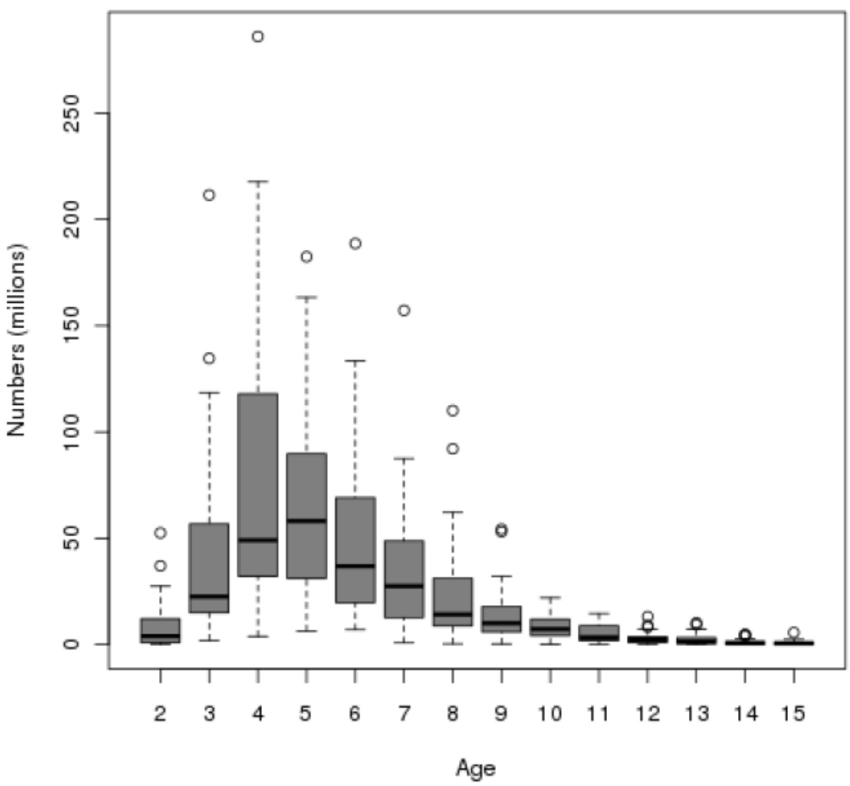


Figure 6. Boxplot of catch-at-age (from WKBENCH 2011).

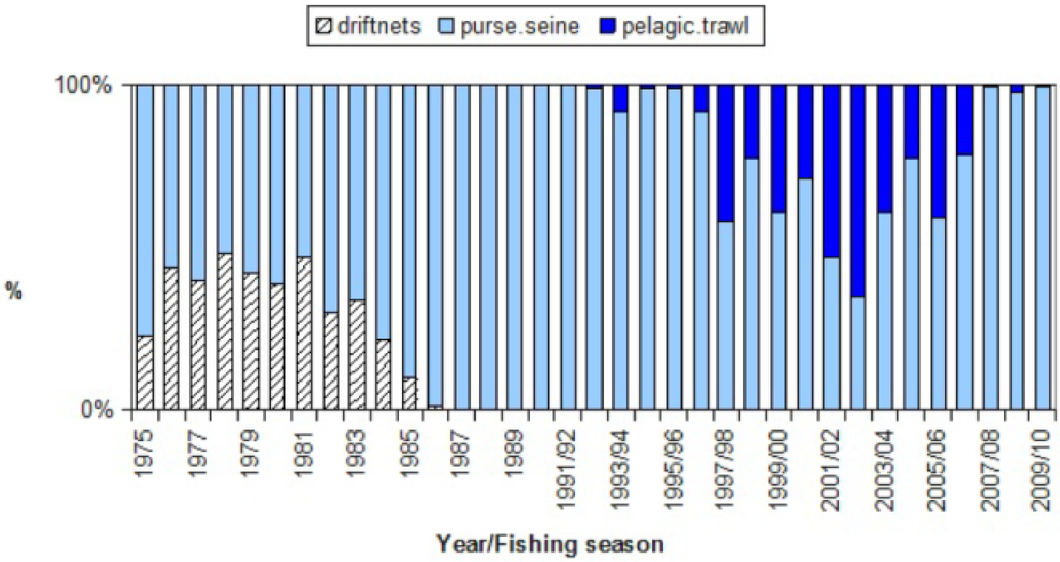


Figure 7. Proportion of landings from different gear (WKBENCH 2011).

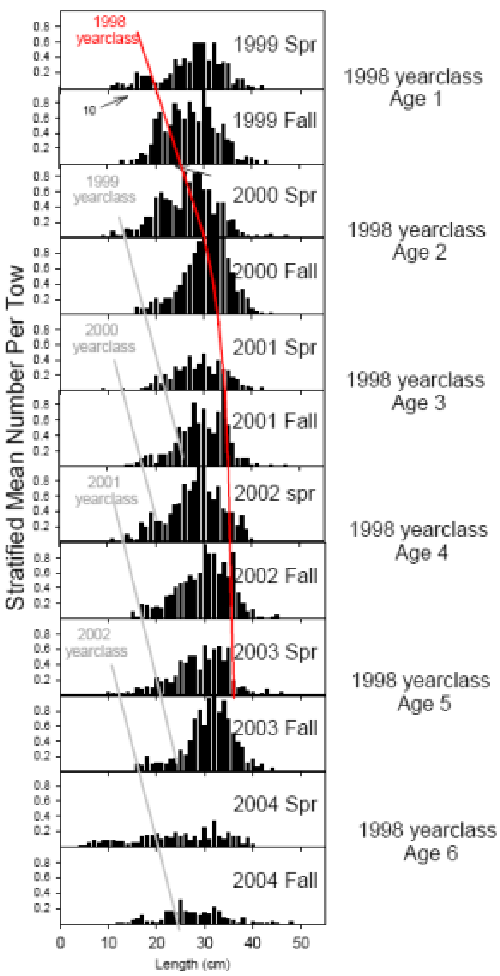


Figure 8. Survey catch at length highlighting a particular cohort (NOAA GARM review).

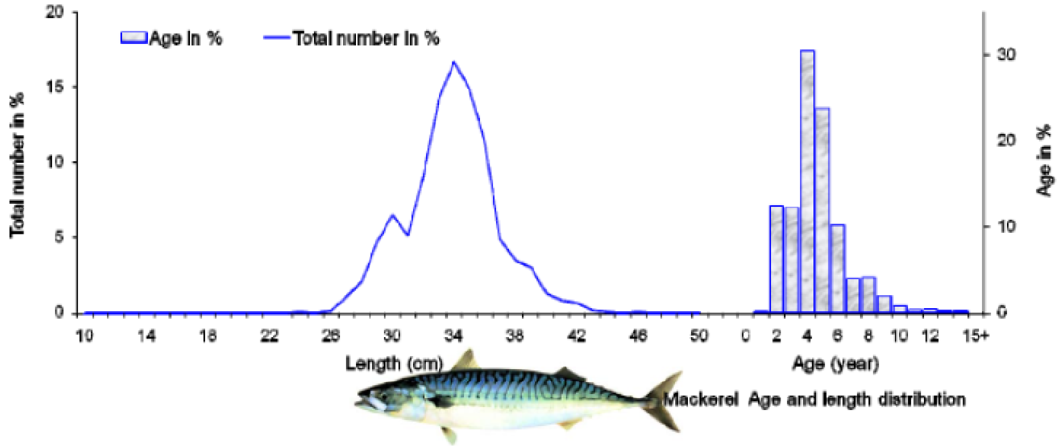


Figure 9. Combining age and length distributions (WGWIDE 2010).

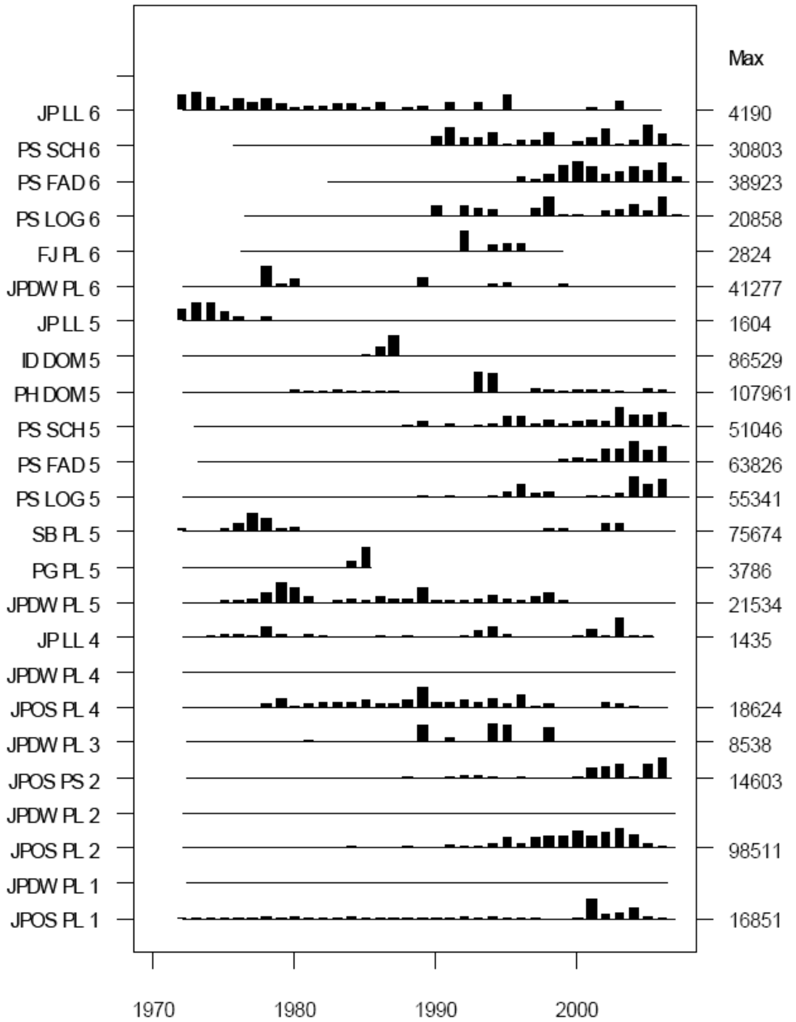


Figure 10. Numbers of length measurements by fishery. Horizontal lines indicate the period of operation of fishery. Each row is proportional to the maximum in that fishery, shown on the right axis (WCPCF skipjack Tuna 2008).

Cpue and surveys

Cpue and survey data can be plotted in exactly the same ways as any other catch data, but there are additional plots that may be useful in examining the data for internal/between survey consistency.

Figure 11 is an extension of the plots of catch, such as Figure 1, to include error bars. Considering the level of error in indices or catch could inform the choice of assessment methodology.

To show age composition, Figure 12 shows each age of the survey as a line, normalized so that it has mean zero and unit standard deviation. On the left hand plot, the x-axis represents years, so that year effects can be seen by consistent patterns in all ages. On the right hand plot, the x-axis represents cohorts, so that cohort effects cause consistently low/high values across each of the lines. This same normalization can be applied to bubble plots to show both age and cohort effects in a single plot.

To check between year consistencies in the survey, CPUE at an age in each year can be compared to the cpue of the same cohort in the following year. An example of this is in Figure 13. Possible extensions of this are shown in Figures 14–17 (where the cpue values are logged). In Figure 14 and Figure 15, cpue at each age is plotted against cpue at each other age in that cohort. Although Figure 15 contains twice as many plots as Figure 14, it contains no additional information, as the plots below the diagonal are repeats (transposed) of those above the diagonal. An alternative use for the below-diagonal space is to display the R^2 values of the correlation, as shown in Figure 16. Figure 17 adds confidence intervals to the correlation between ages, to give an indication of whether the linear trend is statistically significant. Figure 18 shows the consistency between two surveys, effectively treating the Q1 and Q3 survey as a single survey. The time period was split into two periods to address concerns about whether the relative catchability of the two surveys had changed over time.

To examine the structure of each survey, it could initially be plotted in the same way as catch is in Figure 5. Testing for consistent signals across ages from a survey could be done using a plot such as Figure 17. This could also be extended to compare surveys in the case of semi-annual surveys covering the same age ranges, in other cases it would be more difficult to compare the signals from the surveys at the data preparation stage, and this may be done by using diagnostics from the assessment.

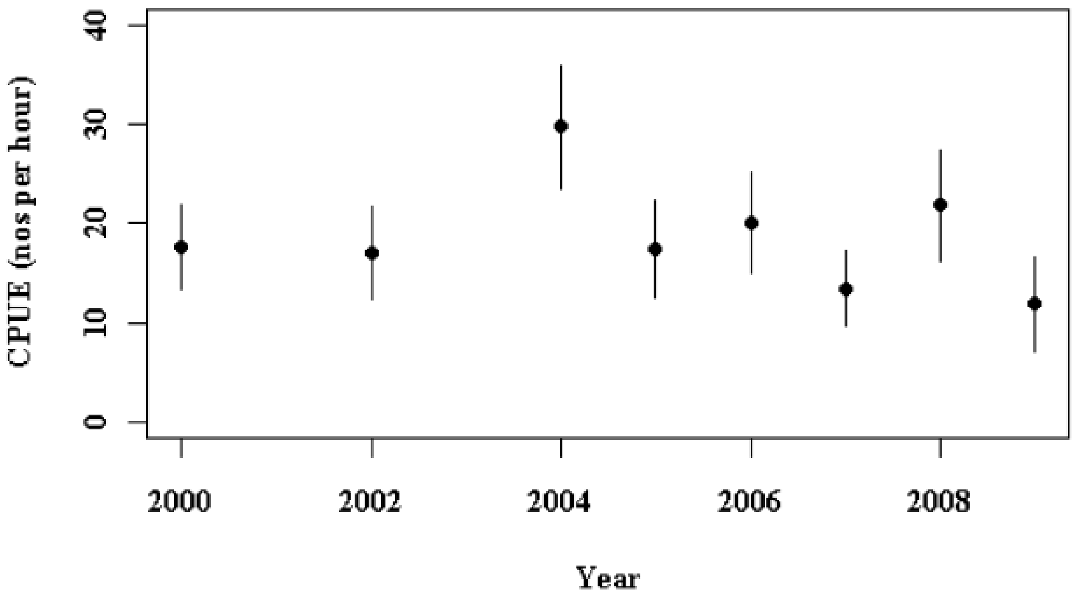


Figure 11. Cpue with error bars from WKDEEP 2010.

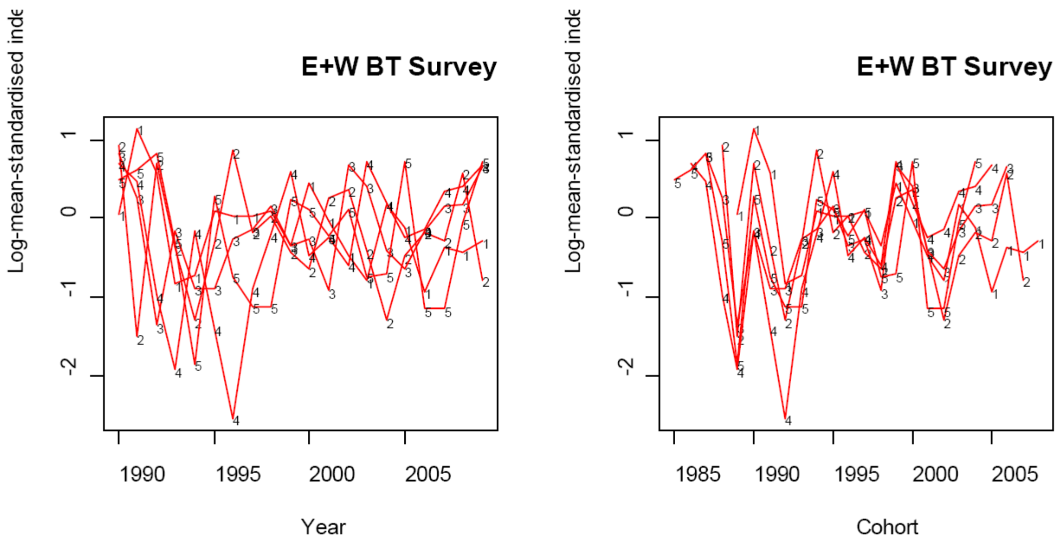


Figure 12. Log mean standardized cpue by year (left) and cohort (right), from WKFLAT 2011.

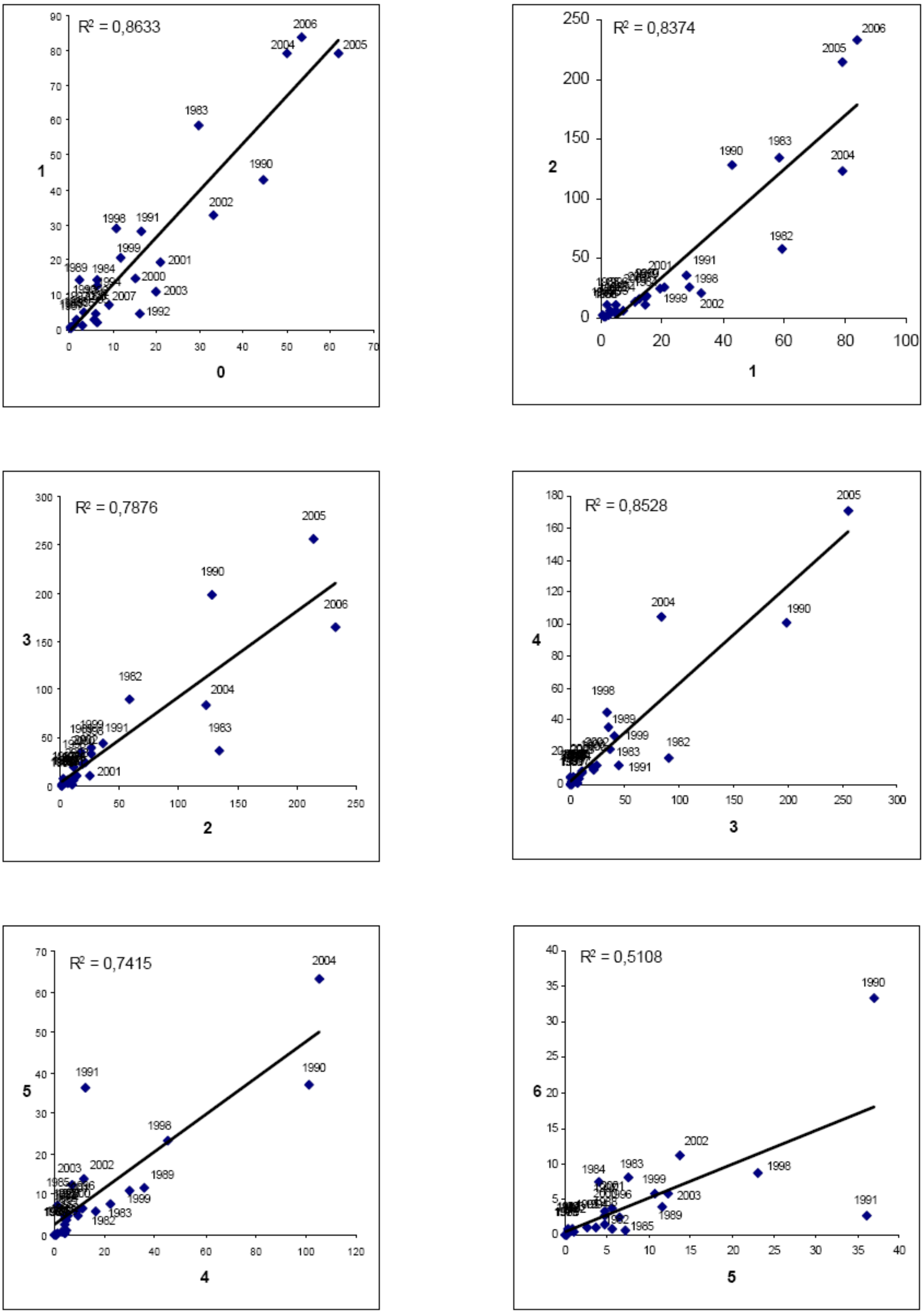


Figure 13. Correlation of consecutive ages in cohort, from WKBENCH 2011.

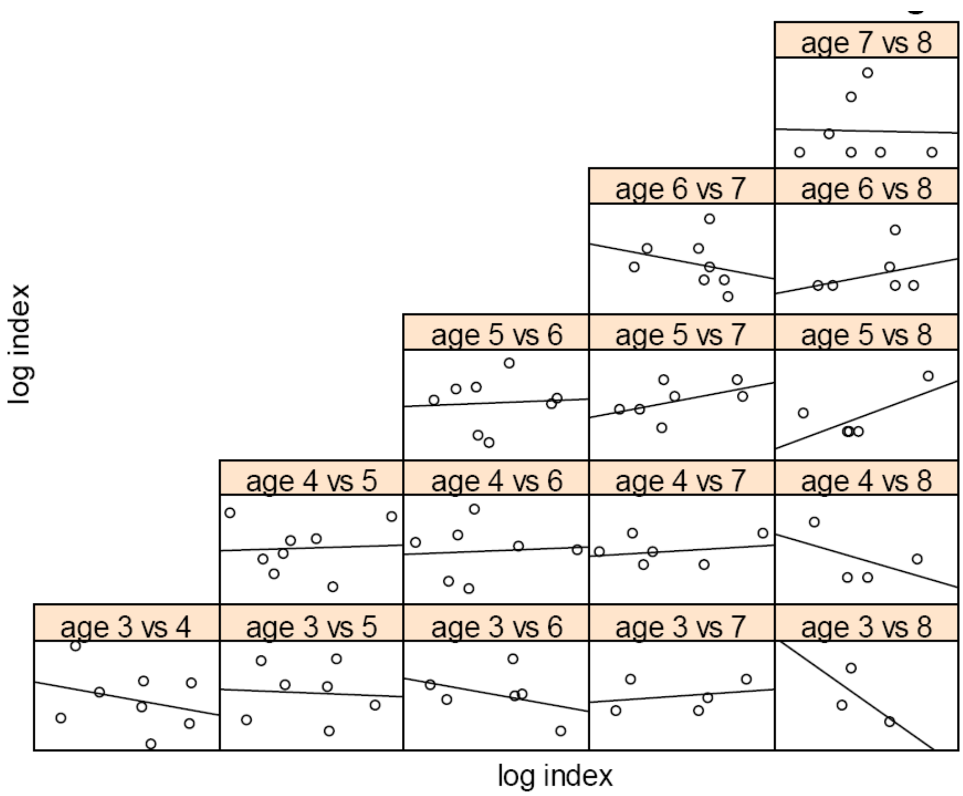


Figure 14. All age correlation of survey index (from WKFLABA).

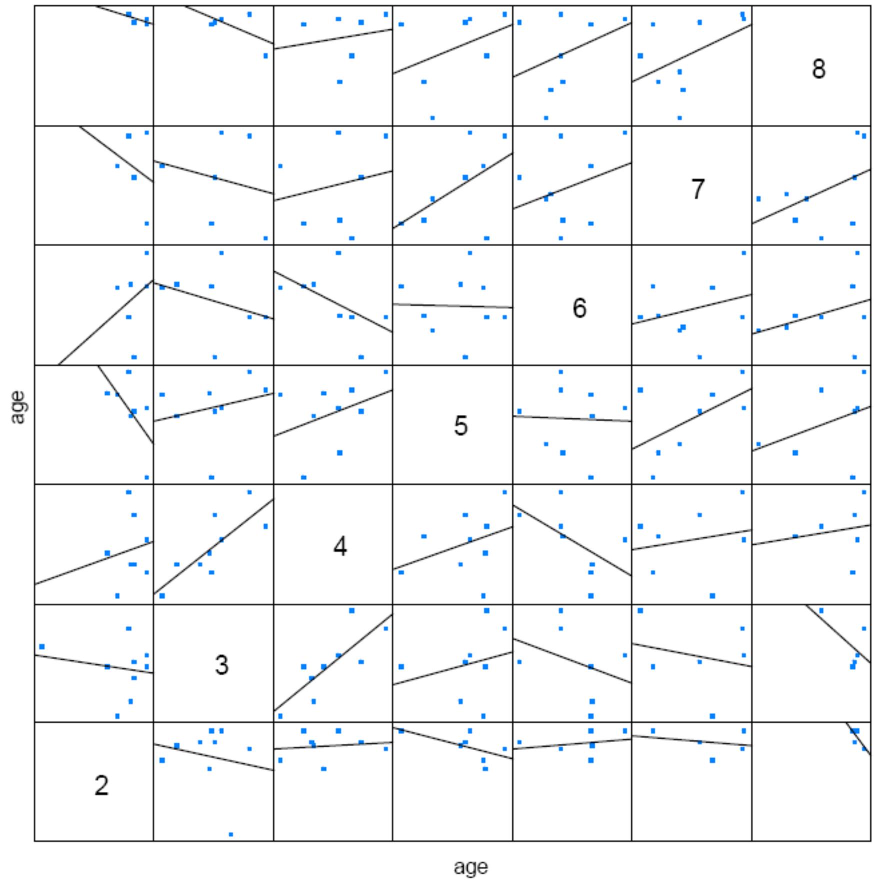


Figure 15. Cpue internal consistency, from WKFLAT 2011.

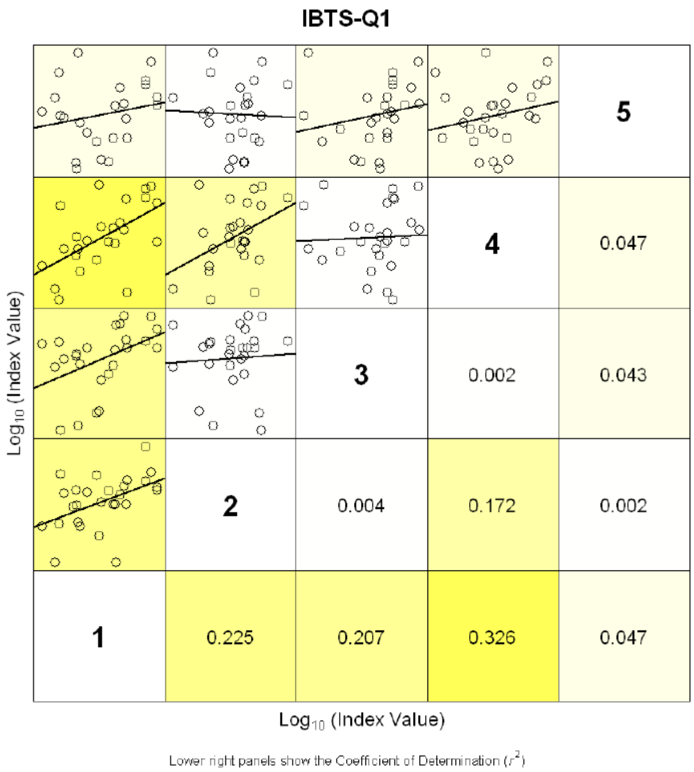


Figure 16. Internal survey consistency with r^2 values (HAWG 2011).

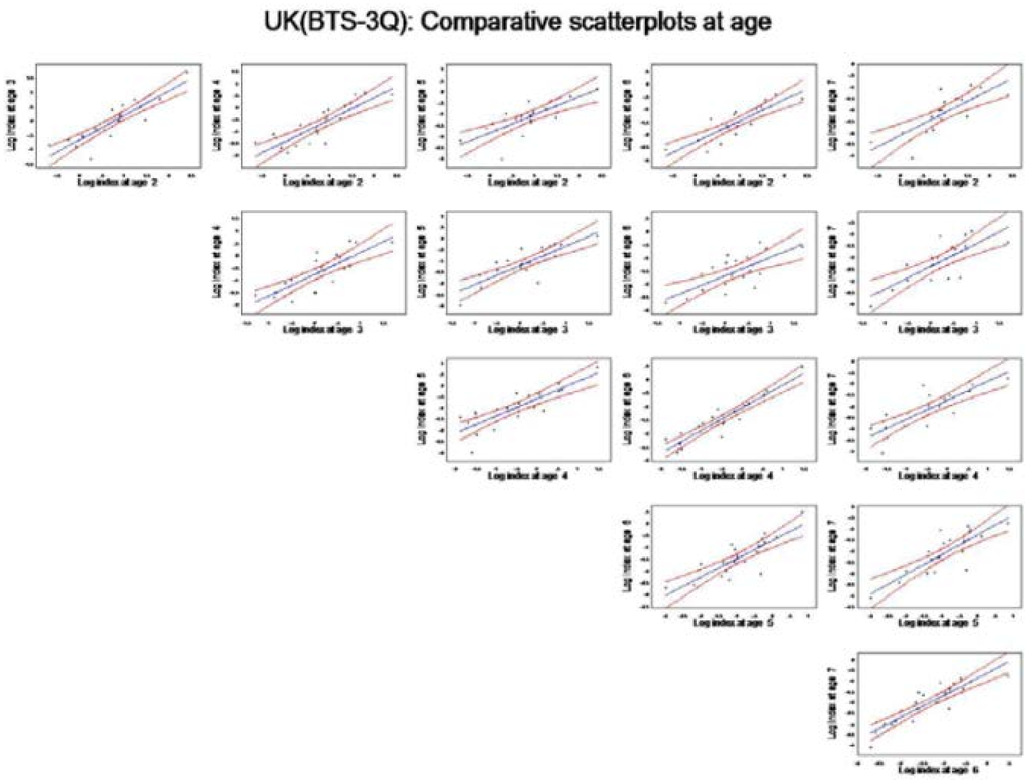


Figure 17. Internal consistency of survey data, with confidence intervals (from WKFLAT 2011).

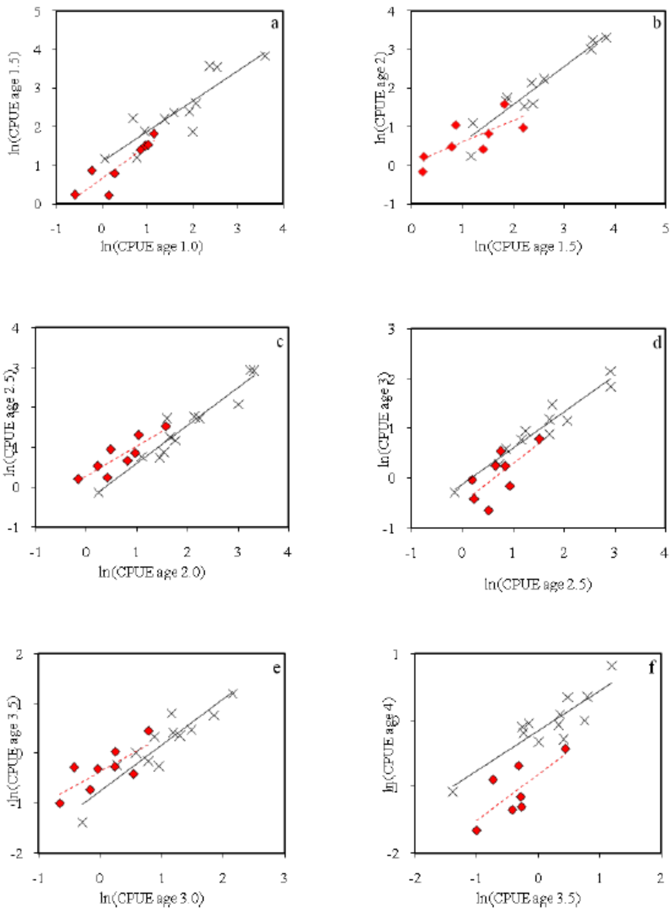


Figure 18. Consistency between Q1 and Q3 surveys. Diamonds indicate most recent years. From WKCOD 2011.

Comparing survey and catch

Figure 19 compares the proportions at age in the catch and the survey, and shows how this changes over time. In most years, the proportion at age 1 is smaller in the catch than the survey, so years where this is not the case, such as 1996 may indicate low recruitment.

Another way to combine survey and catch data are shown in Figure 20, which shows catch (split into landings and discards) on the left axis, and survey index on the right axis (including a rolling average). This presents a good overview of the stock evolution over time, but does not display age structure information.

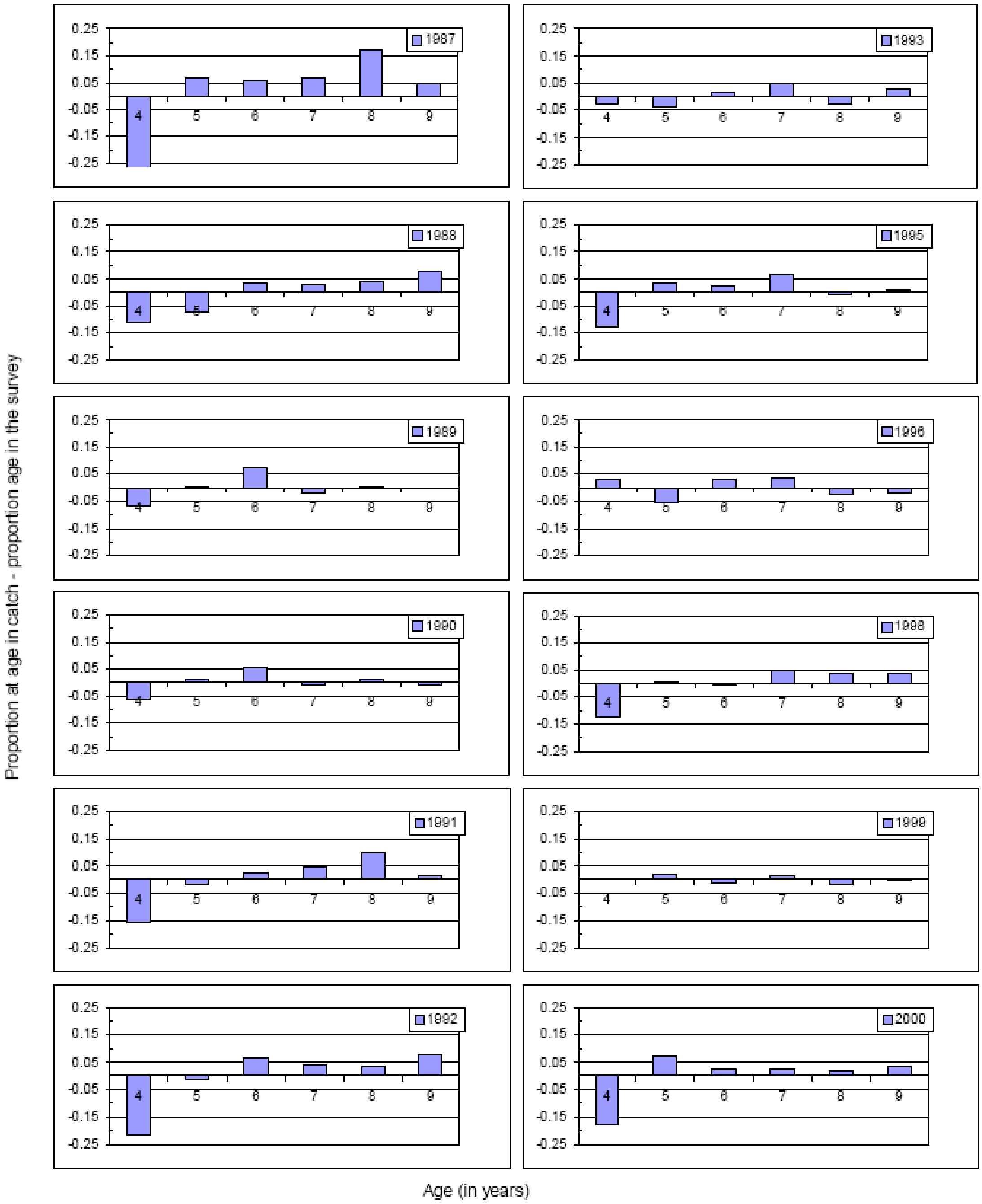


Figure 19. Difference between proportion at age in the catch and in the survey (from WKBENCH 2011).

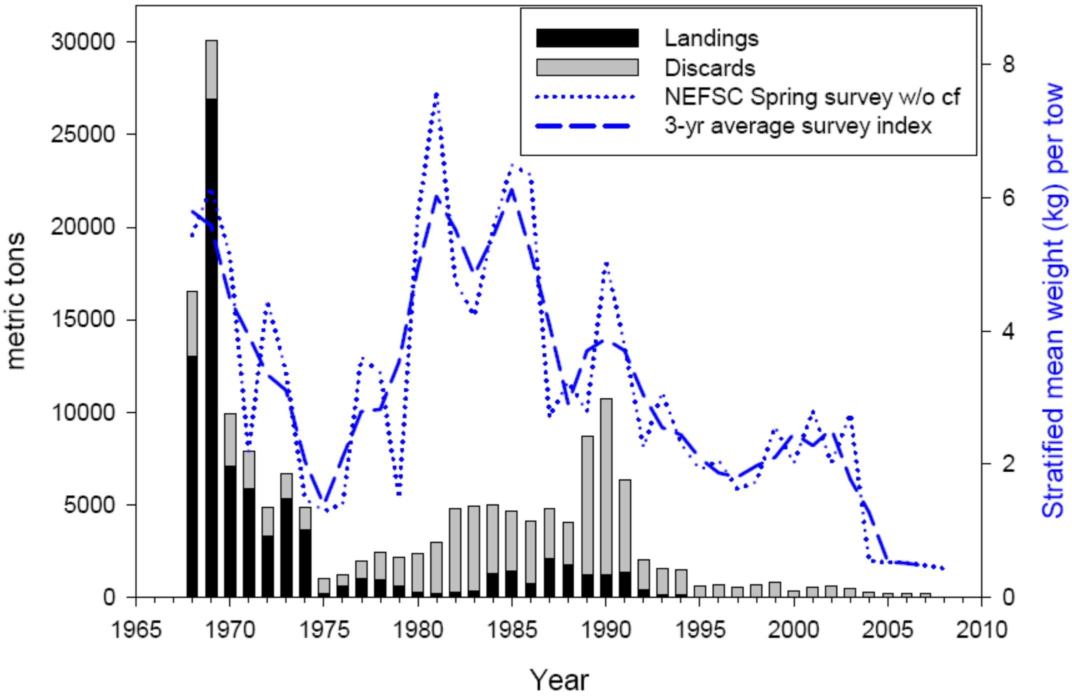


Figure 20. Combining catch and survey time-series (NOAA GARM 2011).

Catch curves and catch ratios

Catch curves plot the ratio of log cpue (or catch) for consecutive years (or other periods) in the same cohort, so that the gradient gives a proxy for total mortality *z* (for ages that are fully recruited to the catch). Figure 21 shows catch curves from a combination of two surveys at six-monthly intervals. The most recent years are coloured differently to highlight that they have a shallower slope, and hence lower mortality than the earlier part of the time-series. An equivalent plot is shown in Figure 22, except that the cohorts have been separated horizontally. The second part of this plot shows the average gradient of each cohort, giving an indication of the trends in total mortality. Careful thought needs to be given to the age range over which the gradient is calculated, so that it represents the fully selected ages as far as is known, so that the strength of individual cohorts does not add variability to the gradient.

Figure 23 shows catch curves based on landings rather than cpue, with each cohort shown in a separate plot. The shading in the background of the plot has an angle corresponding to z=0.4, but is hard to see as the lines are very fine. Figure 24 shows the equivalent of the right hand plot from Figure 22, i.e. the gradient of the catch curves, as a bold line. The plot also shows the equivalent gradient for each pair of consecutive years.

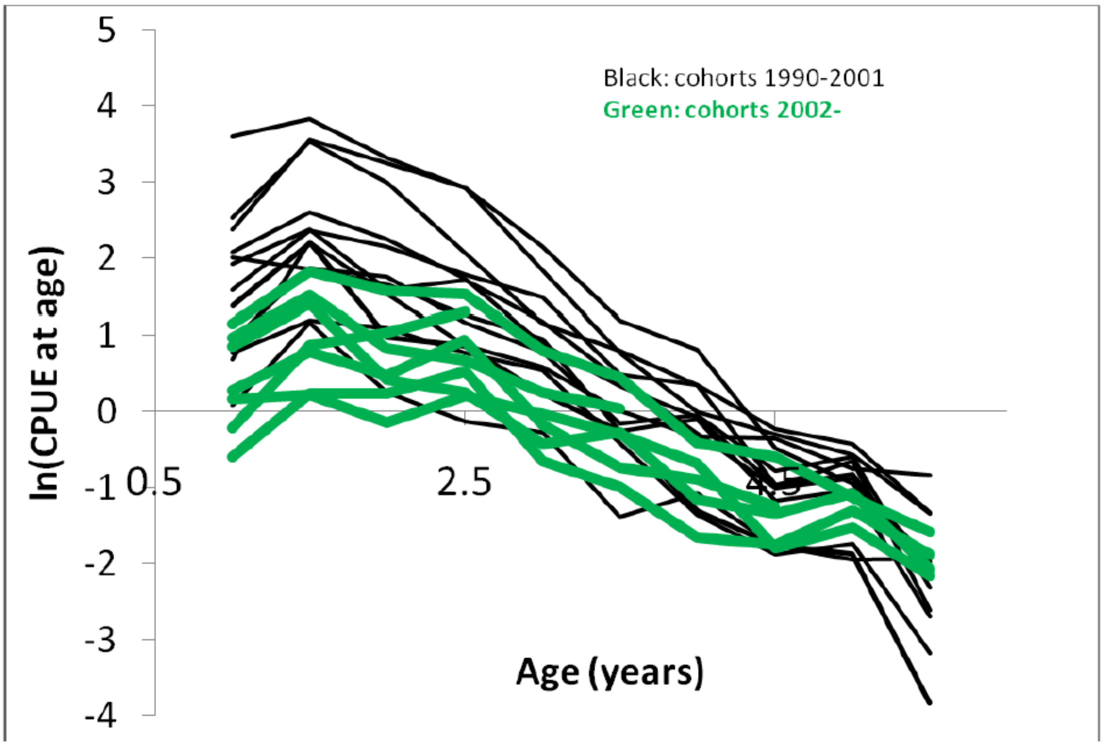


Figure 21. Catch curves by cohort from Survey data (WKCOD 2011).

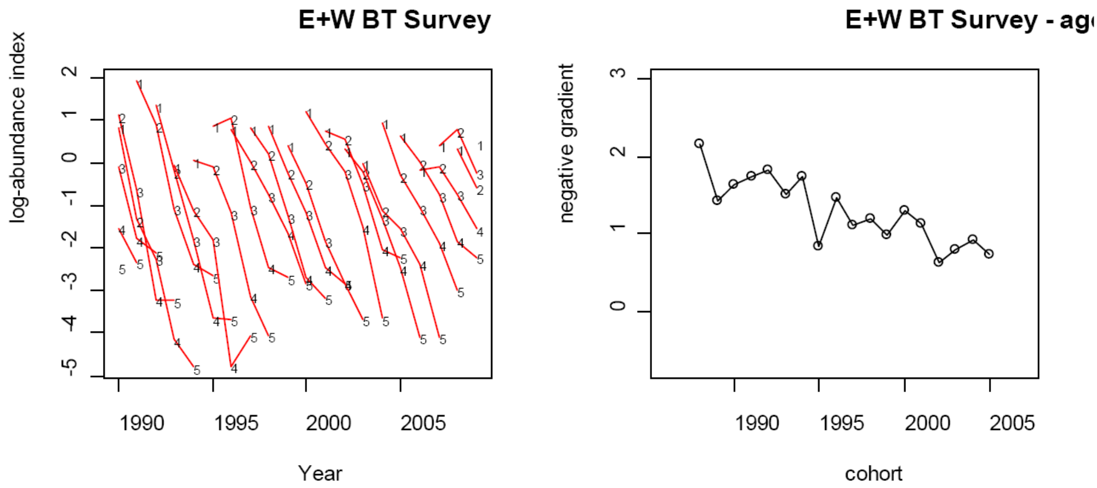


Figure 22. Catch curves and average gradient across cohorts from WKFLAT 2011.

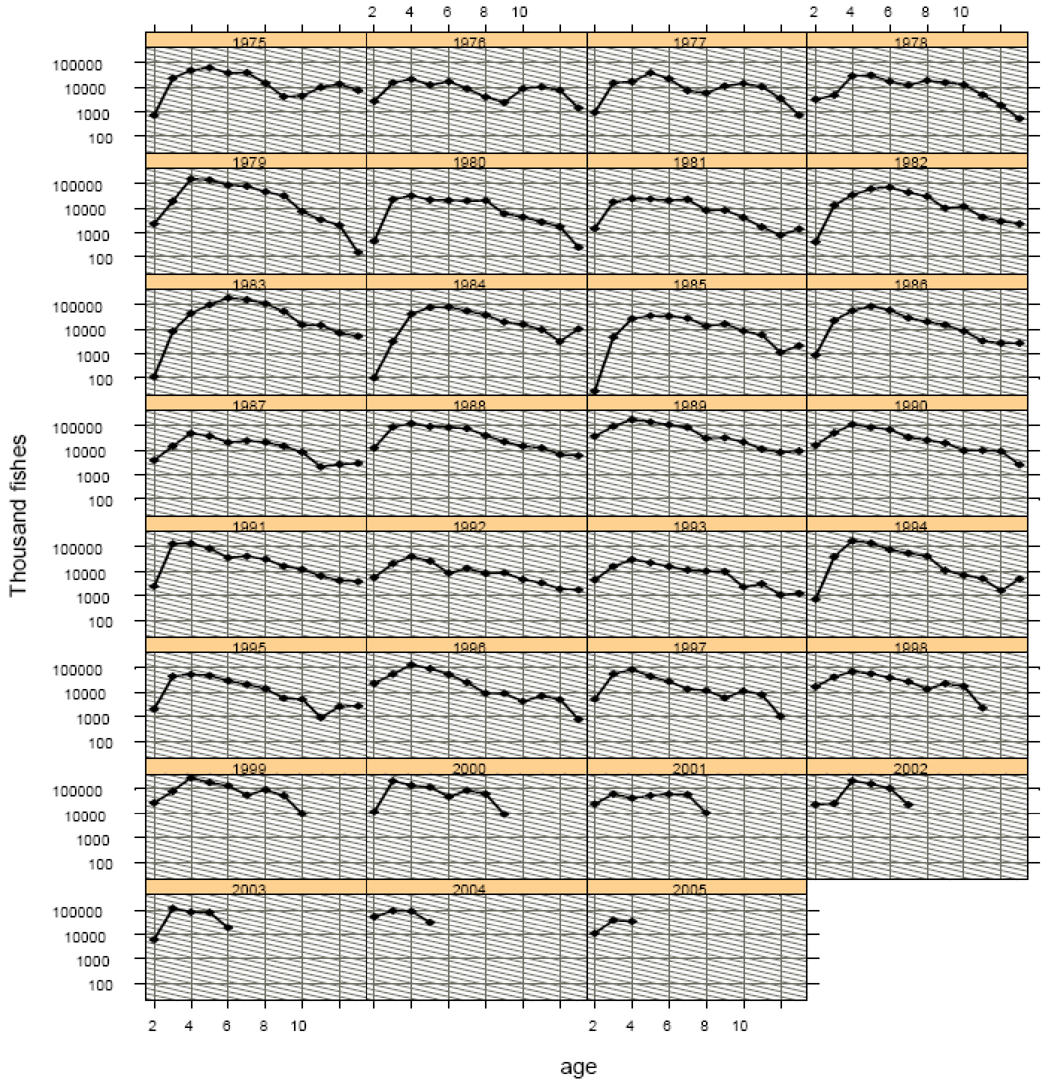


Figure 23. Catch curves for individual cohorts. Diagonal shading indicates Z=0.4 (WKBENCH 2011).

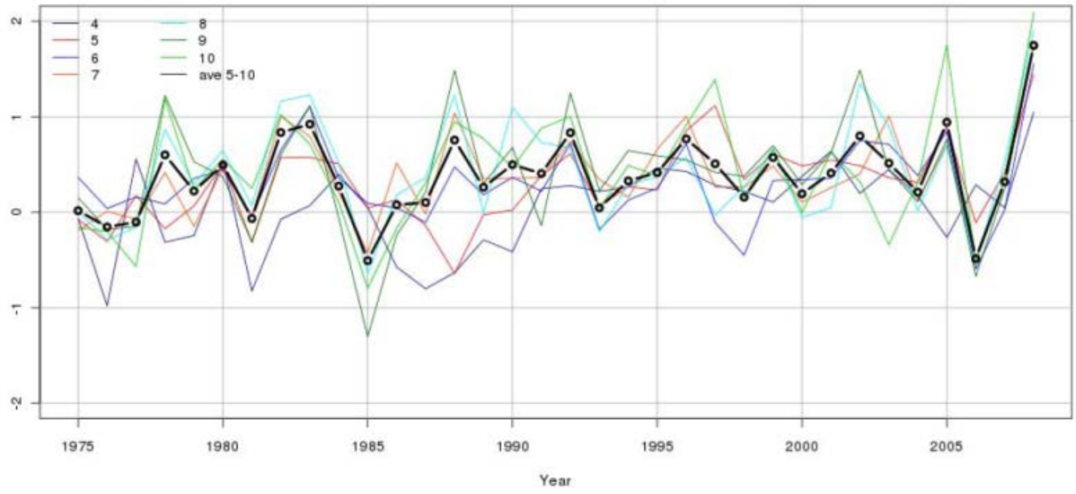


Figure 24. Log catch ratios. Black line indicates average of ages 5–10 (WKBENCH 2011).

Spatial distribution

Spatial data are probably not as widely used as total catch-at-age data, but may be important to address specific issues, such as stock migration. The most basic approach to spatial data are to display catches by landing country or port, such as shown in Figure 25. This sort of data could equally well be displayed as a bar chart without the potential confusion over bubble plot scales mentioned earlier.

The remaining plots in this section cover the locations of catches rather than landing. Figure 26 shows a matrix of maps, one for each combination of age and year. The scale shows the square root of abundance to make it easier to perceive differences at low abundances than would be the case on a linear scale.

Figure 27 shows total catches spatially, using bubbles instead of colours. The key clearly indicates how the size of bubbles relates to the catch weight. An extension of this, shown in Figure 28, is to turn the bubbles into pie charts, in this case indicating the type of gear used.

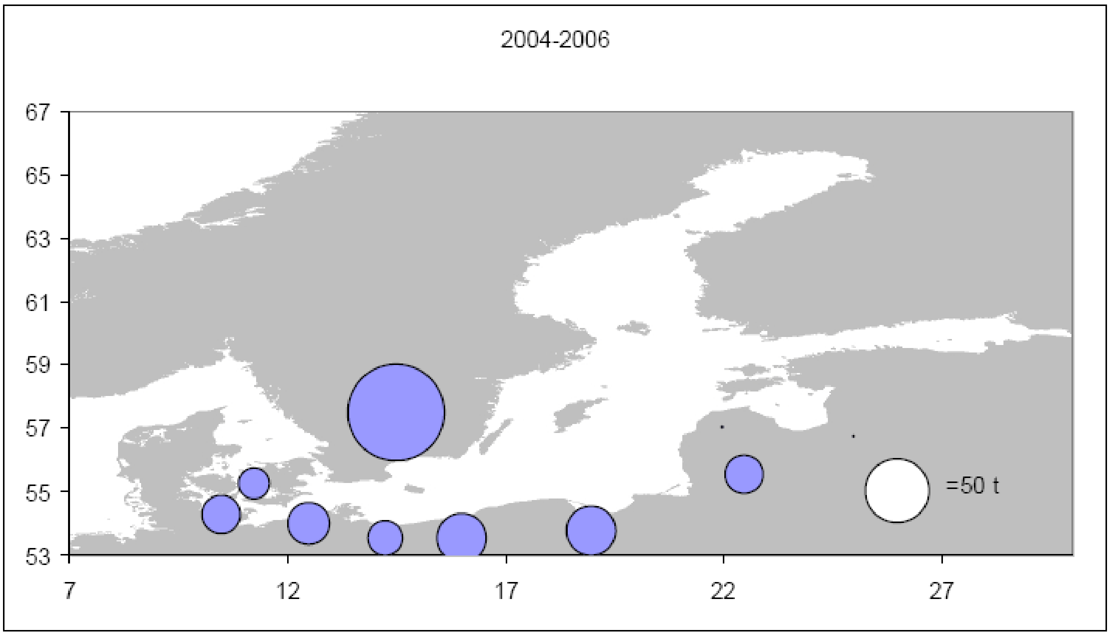


Figure 25. Geographic Distribution of Landings from WKBALTEEL 2010.

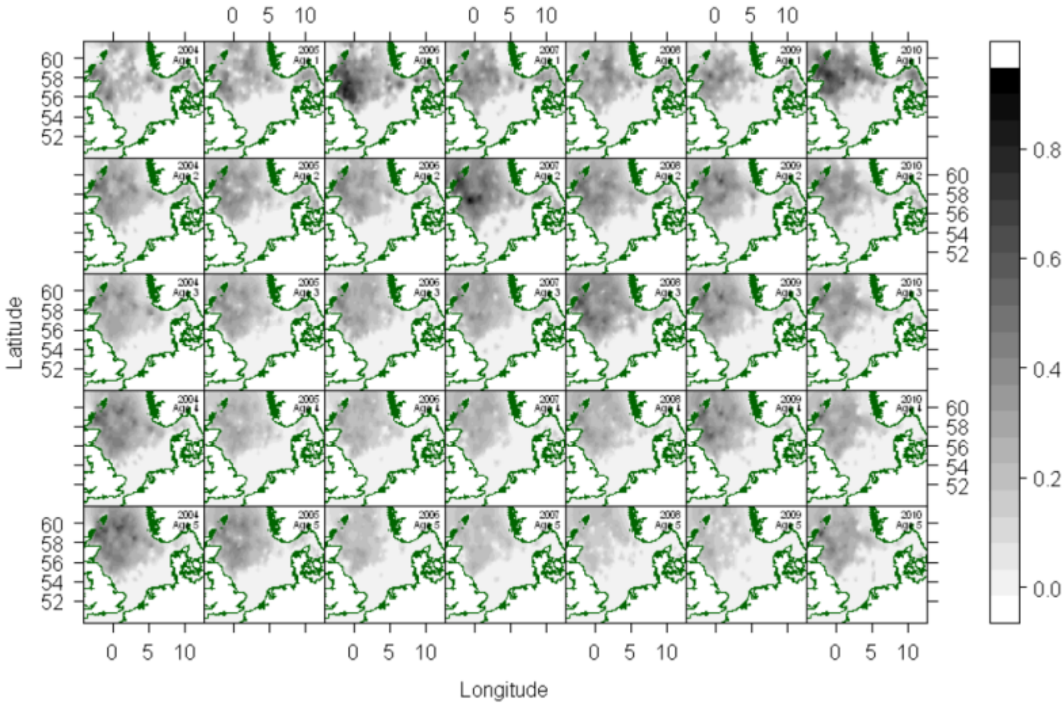


Figure 26. Survey results by age (columns) and year (rows). Plotted values are square root of survey abundance. From WKBENCH 2011.

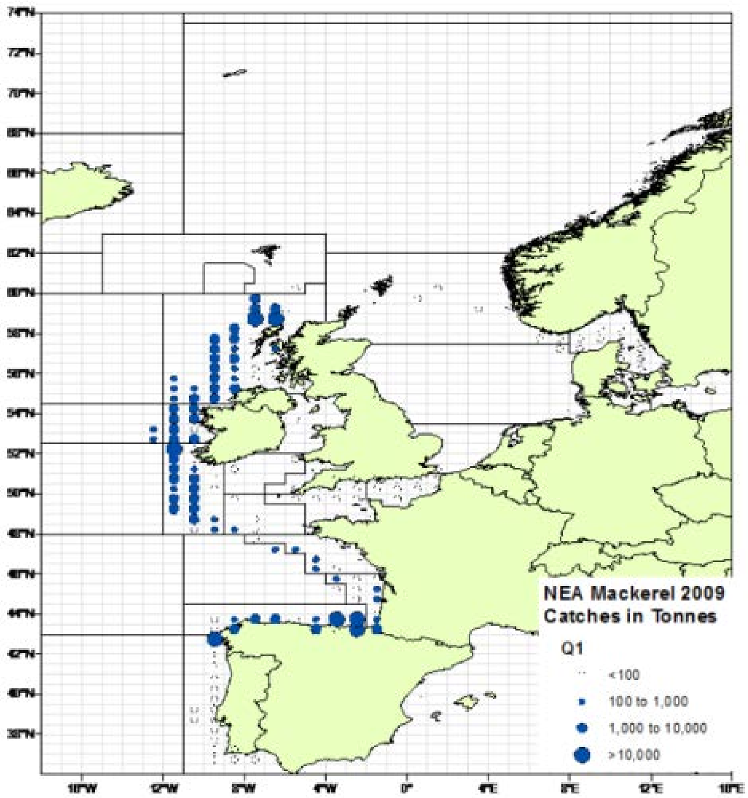


Figure 27. Geographic distribution of Mackerel Catches from WGWIDE 2010 (quarter 1).

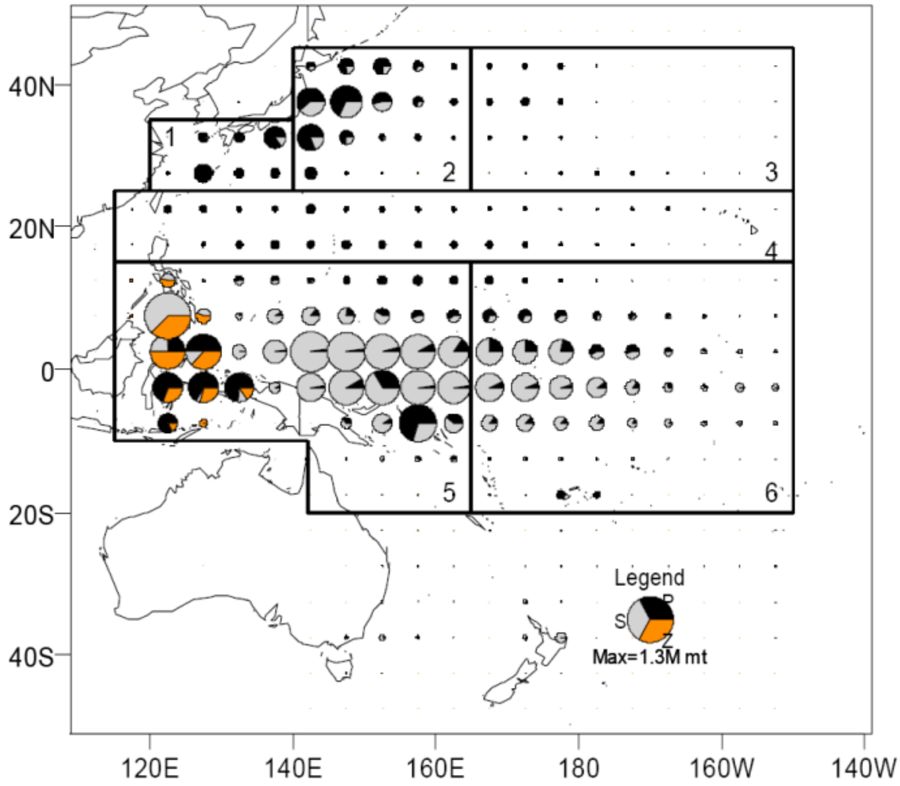


Figure 28. Spatial distribution of catches broken down by gear type (WCPCF skipjack Tuna 2008).

References

Richards, L. J., Schnute, J. T., and Olsen N. 1997. Visualizing catch-age analysis: a case study. Can. J. Fish. Aquat. Sci., 54: 1646–1658.

Skipjack Tuna:

http://www.wcpfc.int/system/files/stock-assessment-specialist-working-group-working-papers/SA-WP-4%20%5BSKJ%20Assessment%5D.pdf

Working Document 2

New runs on Retrospective indices as a measure of input data random error in fish stock assessment.

Antonio Vázquez and Mónica Mandado
Instituto de Investigaciones Marinas, Vigo, Spain

Introduction

During presentation of the work on retrospective indices by Monte Carlo simulation, it was questioned the determining effect of input variability type on the results and conclusions. The original simulation was carried out assuming Normal dispersion for Partial Recruitment (PR) and lognormal dispersion for catch-at-age data (CA) as well survey indices at age (SI; Vázquez and Mandado, 2010). In order to give an explanation to that question, new runs were done assuming Normal dispersion for PR, CA and SI.

One of the objectives of the work is to identify some indices that could be used to measure disagreement in retrospective analysis. The desirable properties of a good retrospective index, to be useful as indicator of problems in the input data or in the whole analysis, are:

1) It should be zero when input data inaccuracy and VPA model disagreement are both null.
2) It should be a positive magnitude.
3) It should increase when input data inaccuracy increased and when VPA model disagreement also increased, preferably with a linear relationship.
4) It should have the lowest dispersion for a given error of input data.
5) It should be correlated with some inaccuracy index.

The first condition was satisfied by all tested indices. It was checked systematically as a way to verify the routines.

The second condition was introduced because the sign of output is irrelevant when the effects of random variability were only tested. This condition eliminates ϱ indices (arithmetic mean deviation) from further consideration, even they remain a useful tool to detect retrospective patterns.

Even the ϱ indices are considered inappropriate for these analyses, question was rise on the reason why it resulted in a positive mean. It was proposed that it was due to the lognormal dispersion used in the simulation. The implications of the lognormal assumption in the whole analysis is considered here and compared with results based on an alternative Normal dispersion assumption.

Methods

Random distribution generators were review:

The Normal distribution generator was the Marsaglia-Bray algorithm (Marsaglia and Bray 1964) modified by Shonkwiler and Mendivil (2009).

N(0, 1) = {v1 * sqrt (– 2.0 * log(S) / S)}

S = $v1^2 + v2^2$ (with the condition: S<1)

v1 = rand() * 2.0 – 1.0 (rand() = Uniform [0, 1])

v2 = rand() * 2.0 – 1.0
So: {x} = N(0, 1) => {x*sd + m} = N(m, sd)

The lognormal distribution is produced from a Normal one:

Let it be: {x} = N(0, σ)
Then {EXP(x)} = logN(median=1, s), $s^2 \sim EXP(\sigma^2) - 1$
{m EXP(x)} = logN(median=m, sd=m·s)
$sd^2 \sim m^2 [EXP(\sigma^2) - 1]$
$\sigma^2 \sim \ln(1 + sd^2/m^2)$
So: {x} = N(0, $\sigma^2=\ln(1+sd^2/m^2)$) => {m·EXP(x)} ~ logN(median=m, sd)

These random Uniform, random Normal, and random lognormal generators were checked, and then used to produce partial recruitment, catch-at-age and survey indices at age with the following standard deviations:

MAIN SOURCE OF VARIABILITY	CONDITIONING CONSTANTS AND SD USED		
	sdPR	**sdCA**	**sdSI**
Partial recruitment	0.0001, 0.1, 0.2, 0.5, 1.0	0.2	0.3
Catch-at-age	0.2	0.0001, 0.1, 0.2, 0.5, 1.0	0.3
Survey indices at age	0.2	0.2	0.0001, 0.1, 0.2, 0.5, 1.0
Same sd in each constant	0.0001, 0.1, 0.2, 0.5, 1.0	same as sdPR	same as sdPR

Catches-at-age and survey indices at age generated under this conditions were VPA input data for analyse. Each trial run contained 1000 iterations, with one randomly generated input dataset each, being independently analysed with two ADAPT formulations and XSA. *Retrospective indices* and *inaccuracy indices* (*bias indices* in the original paper) were calculated from results in each of three analyses.

Results

Only results from the above table under "Same sd in each constant" were presented in Figures 1 and 2 for illustrating purposes. Each one of the five standard deviation test values was applied simultaneously to the three sources of variability in that case. These figures show mean retrospective indices calculated with the standard deviations pointed out in the above. Figure 1 was done as the original paper: Normal dispersion for PR and lognormal dispersion for catch-at-age data and survey indices at age. Figure 2 was done assuming Normal dispersion in all cases. These two figures are presented to illustrate the common facts observed in all results, which are:

- Indices based on ADAPT with 9 or 10 parameters are quite similar, and behave as expected, increasing with higher dispersion of input data.
- Indices based on XSA do not start at (0,0) because shrinkage on final year and oldest age survivor were set, and it implies some disagreement in the VPA model. It could be avoided, but it was maintained to illustrate that source of inaccuracy which is not quantified.
- No substantial differences were observed in retrospective indices behaviour if they were produced with lognormal or Normal dispersion of input data. This was one of the points to analyse in this working paper.

Figures 3 and 4 shows mean inaccuracy indices corresponding to Figures 1 and 2 respectively.

- Only ϱ inaccuracy indices are affected by the kind of dispersion in input data. Mean ϱ indices became below zero, particularly with high dispersion.

- Mean σ and π inaccuracy indices seems to have quite similar behaviour between figures, indicating they are not dependent of the dispersion type for input data. This was another main point to analyse in this paper.

Figure 5 is based on σ1 retrospective index vs σSSB3* inaccuracy index individual values, instead of using means values as in previous figures. This figure is the key in establishing a relationship between both indices. Distribution of points is quite similar in both cases, which confirm its independence from the dispersion type for input data.

Points of Figure 5 proceed from the ADAPT with 9 parameters. An analysis is still missing on the dependence of this relationship from the VPA method used. As it was observed in Figures 1 to 4, ADAPT and XSA do not produce the same retrospective and inaccuracy indices.

Discussion

The results indicate the minor effect of the type of variability on input data, being Normal or lognormal dispersed. The whole VPA analysis behaves as a complex black box, where variability of input signal has minor effects on the type of variability of outputs.

Even the meaning of σ1 and σSSB3* indices is not explained here, but only in the original document, they are only selected to illustrate the possibility of using a retrospective index (σ1) as indicator of inaccuracy (σSSB3*). This selected pair of indices requires further review to identify the more efficient one.

References

Marsaglia, G. and T. A. Bray – 1964. A Convenient Method for Generating Normal Variables. *SIAM Review*, Vol. 6, No. 3: 260–264.

Shonkwiler, R.W. and F. Mendivil – 2009. Explorations in Monte Carlo methods. *Springer*, 243 pp

Vázquez, A. and M. Mandado – 1010. Random Retrospective Pattern in Fish Stock Assessment. *NAFO SCR Doc.* 10/36.

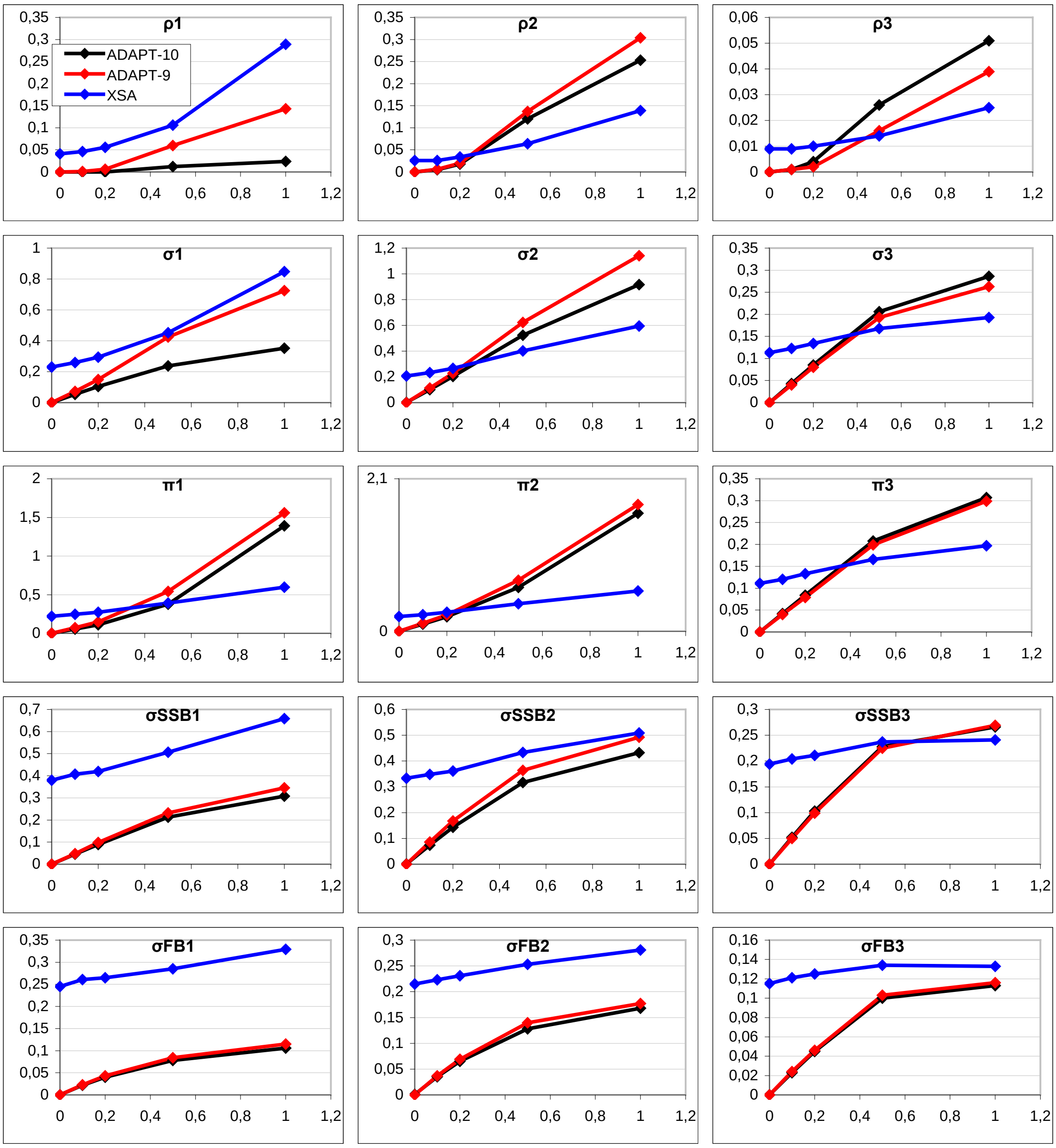


Figure 1. Retrospective indices under Normal dispersion for PR and lognormal dispersion for catch-at-age data and survey indices at age.

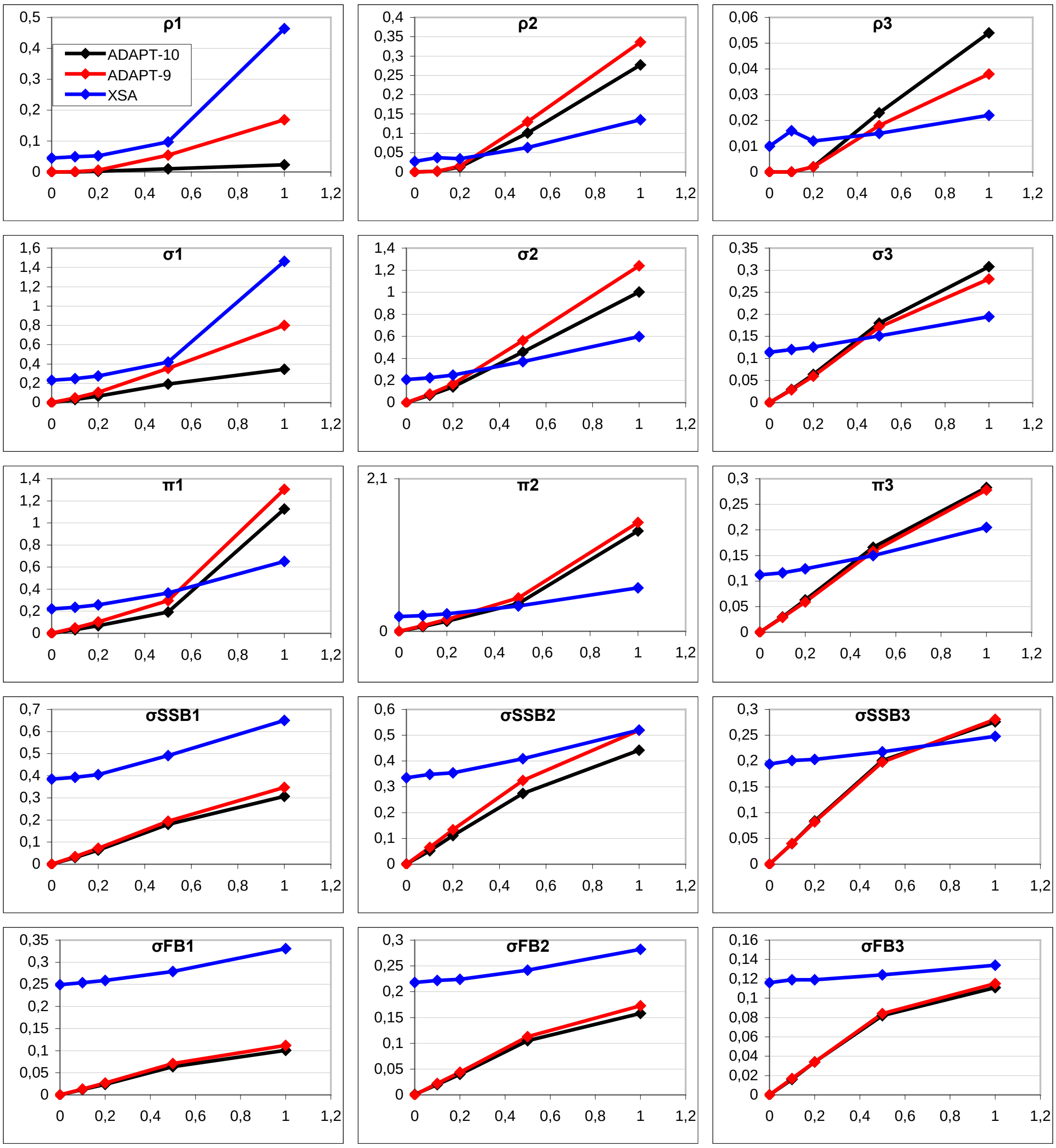


Figure 2. Retrospective indices under Normal dispersion for the three sources of variability: PR, catch-at-age data and survey indices at age.

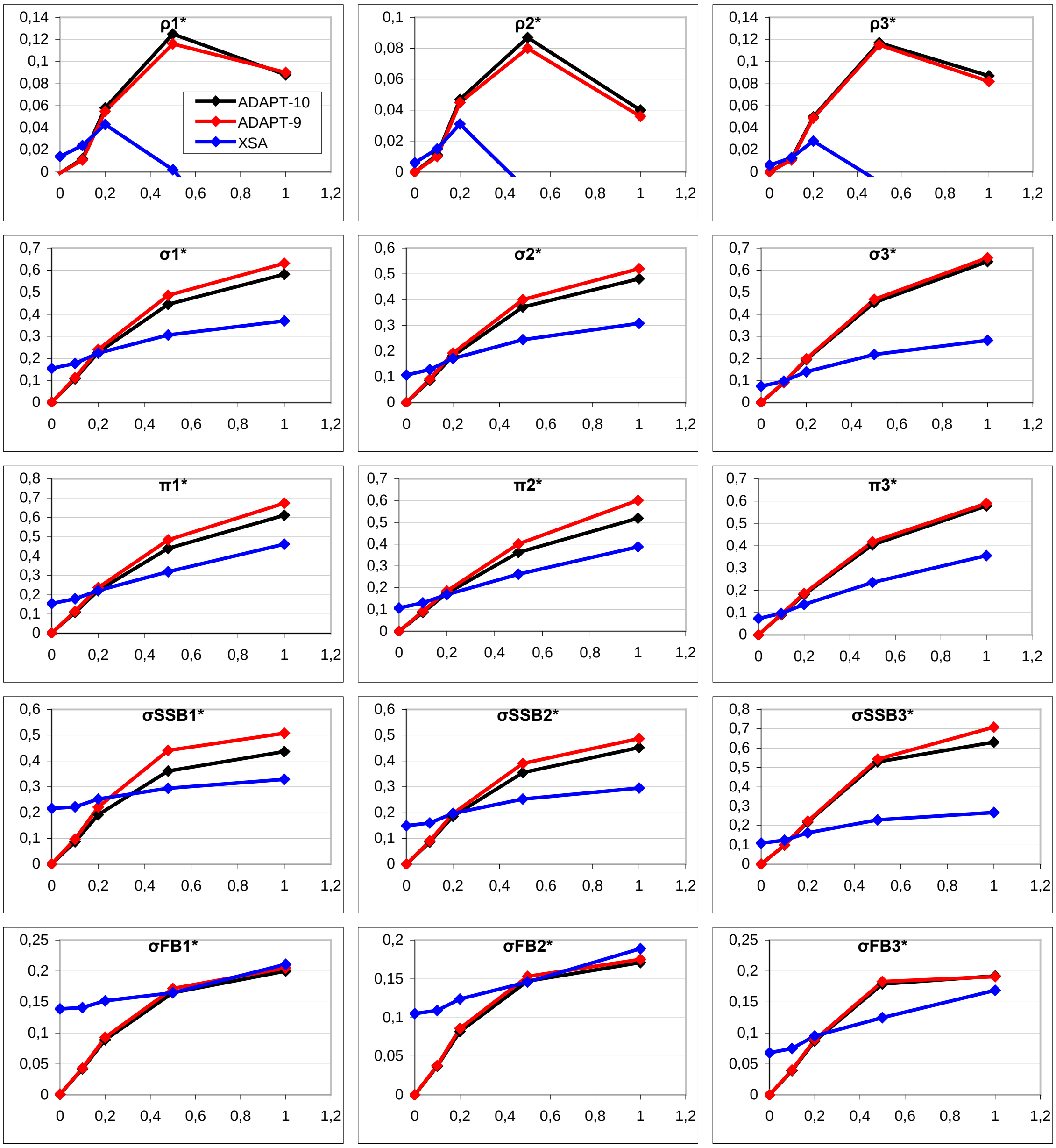


Figure 3. Inaccuracy indices under Normal dispersion for PR and lognormal dispersion for catch-at-age data and survey indices at age. Same data as in Figure 1.

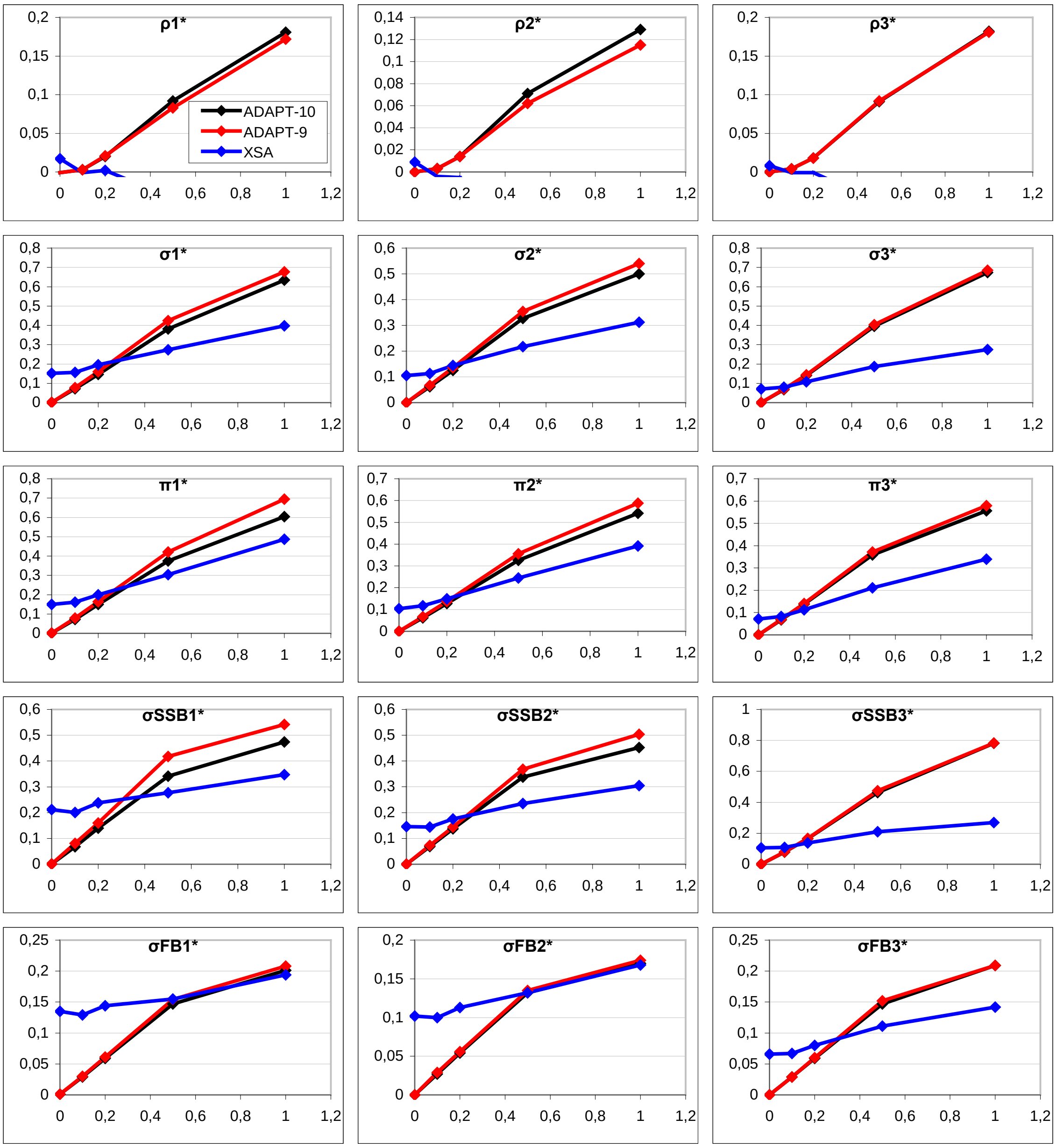
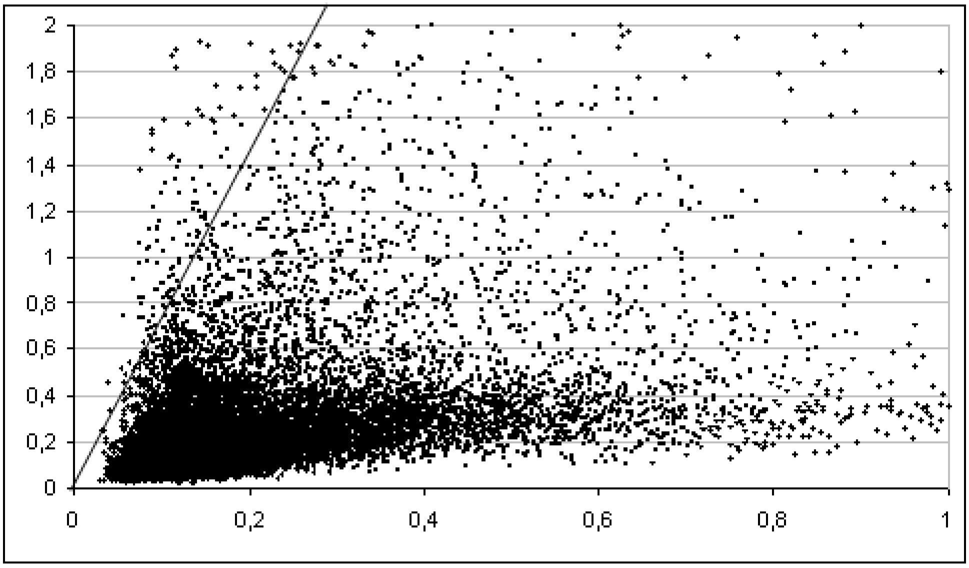


Figure 4. Inaccuracy indices under Normal dispersion for the three sources of variability: PR, catch-at-age data and survey indices at age. Same data as in Figure 2.



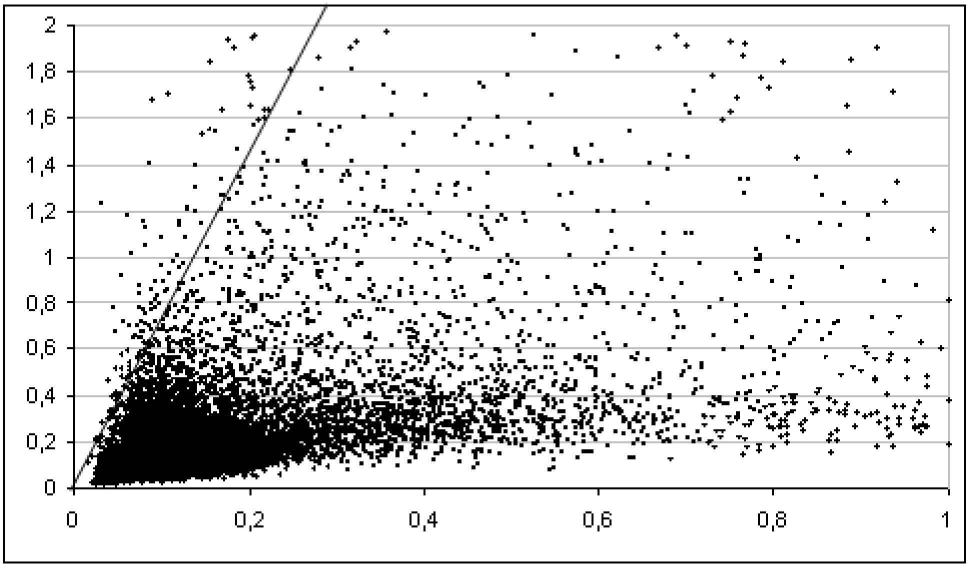


Figure 5. Plot relationship between σ1 retrospective and σSSB3* inaccuracy indices calculated with all cited sources of variability, using the ADAPT with 9 parameters. The upper figure was based on lognormal deviations. The lower figure was based on Normal deviations.

Working Document 3

Some reasons generated the retrospective bias in the stock assessment models

Pavel Gasyukov and Svetlana Kasatkina

At present the Extended Survival Analysis (XSA), developed by Shepherd (1999) is the principle method of commercial fish stocks assessment in the Baltic Sea. Abundance indices by years and age groups represent the most important input data for this method. However, the objective function of the "classic" XSA version considerably differs from the function used in ICES software and applied by ICES working groups in the Baltic commercial fish stocks assessment (Darby, Flatman, 1994). The difference is in that in ICES software the variance of abundance indices, as the indication of these indices accuracy is assumed to be a constant value by years for each age group.

This assumption is not based on any real estimates, but is stipulated by the difficulties in assessment of variance being the function of several influencing factors associated with the surveys.

In the recent years, the methods of simulation for determination of abundance indices statistical characteristics (mean values, variances, standard errors, variation coefficients and confidential intervals) obtained on the basis of data from the Baltic international surveys have been developed at AtlantNIRO (Kasatkina, Gasyukov, 2006). These methods allowed to verify practically the assumption adopted in ICES software for XSA implementation that abundance indices dispersion is a constant value by years in each age group.

The method of statistical characteristics assessment for sprat and herring biomass and abundance indices obtained from the Baltic International Acoustic Surveys (BIAS) data were developed on the basis of simulation using the Monte Carlo method (Kasatkina, Gasyukov, 2006, 2009).

The spatial variability of the acoustic index NASC ($m^2/n.mile^2$), the spatial variability of fish species composition and length structure in the study area and uncertainty of the target strength estimates were considered to constitute the basic sources of uncertainty in BIAS surveys. The4 developed simulation model allows to estimate contribution of each uncertainty source and their total effect. The effect of each uncertainty source was simulated using the bootstrap procedure (Efron, 1988; Efron, Tibshirani 1993): the parametric bootstrap was used in the target strength simulation; the bootstrap with application of empiric distribution functions was used in simulation of variability of the acoustic index NASC, species composition and length structure of fish.

Using the developed simulation method, the statistical characteristics of fish abundance indices were estimated on the basis of BIAS 2004–2006 data, obtained by surveys participants (Poland, Germany, Latvia, Lithuania, Sweden, Estonia and Russia). The input data for simulating included the following:

- Regression equation of the target strength accompanied with statistical characteristics of its parameters;
- NASC estimates ($m^2/n.mile^2$);
- Length frequencies of all fish species, recalculated by the total catch, for each trawl station;
- Age keys for herring and sprat;

- Number of replications in simulation (500 replications).

Some statistics of fish abundance indices are presented on Figure 1.

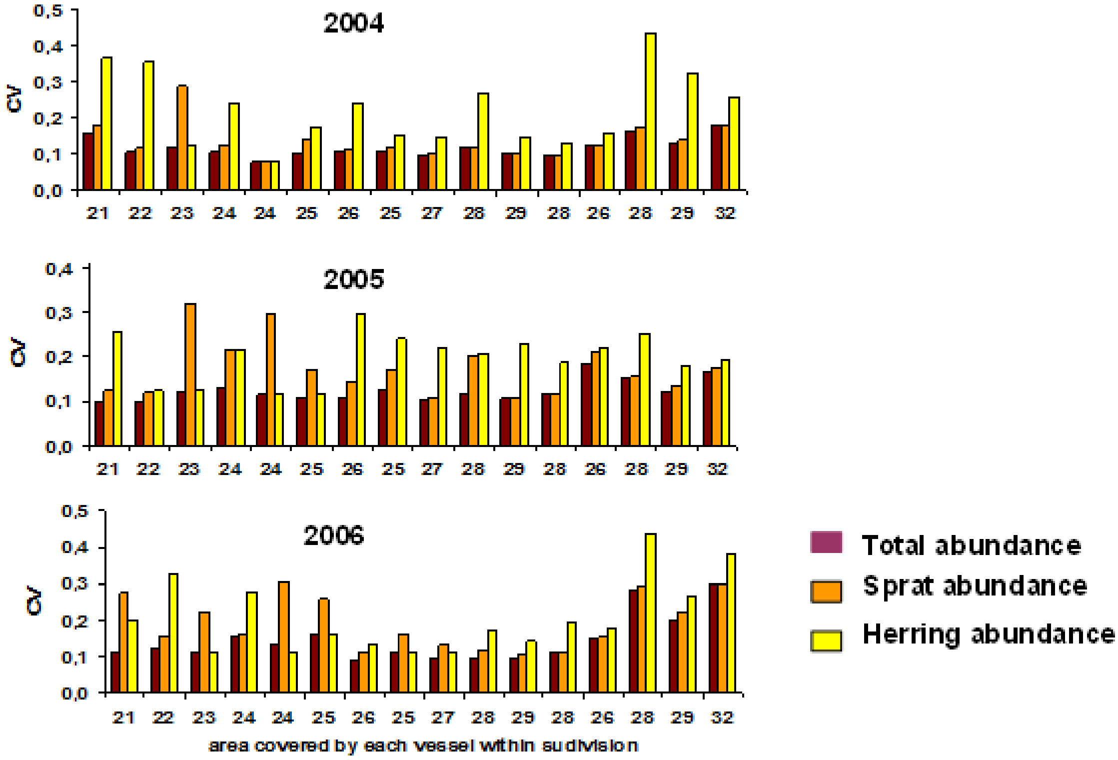


Figure 1. Coefficients of Variation from BIAS surveys 2004–2006 as example. Estimates of herring and sprat abundance and total fish abundance obtained within each stratum (21–32 strata) by some countries have different accuracy.

It was also revealed that abundance indices are characterized by variances which dependent on value of abundance indices themselves, therefore variance is not constant value by years for each age groups.

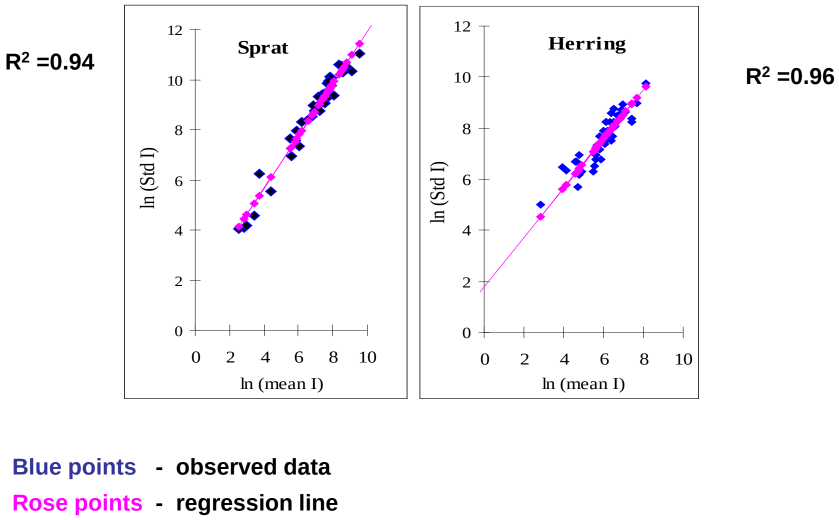


Figure 2. Relationship between logarithm of the abundance index standard deviation and logarithm of the mean abundance index

The relationship between abundance indices variance and indices value likely reflects the pattern of the Baltic *Clupeids* spatial distribution. Therefore, it seems that the stock assessment model accounting for variability *of abundance indices variance* will describe the Baltic fish dynamics more realistically.

The task of accounting for accuracy of fish abundance indices in stock assessment was solved by the way of elaboration of a new XSA version. The following regression equation is one of the principle equations in XSA.

$$I'_{(y,a,f)} = q_{a,f} \cdot N^{\gamma_{a,f}}{}_{VPA(y,a)} \qquad (1)$$

where $I'_{(y,a,f)}$ - abundance index in the year *y* for the age group *a,* and fishing fleet *f,* recalculated to the year beginning;

$N_{VPA(y,a)}$ - abundance estimate obtained with VPA,

$q_{a,f}$ – catchability coefficient, $\gamma_{a,f}$ - exponent.

Indeed the application of equation (1) in XSA gives rise to the idea of the constant variance of abundance indices by years, since this equation assumes that all parameters in the left part of the equation are independent and similarly distributed random values. This equation (1) is transformed into the regression equation for the young age groups, including recruitment:

$$\ln N_{vpa}(y,a,f) = \frac{1}{\gamma_{a,f}} \cdot \ln \mathrm{I}'(y,a,f) - \frac{q_{a,f}}{\gamma_{a,f}} \tag{2}$$

and allows to obtain the abundance estimates $N_{est}(y,a,f)$, used in assessment of survived fish abundance at the end of the terminal year, and for the older age group by years and by fleets (index *f*).

The equation (1) is also transformed into the equation (3) for estimation of the inverse value of the catchability coefficient for the rest age groups:

$$\ln N_{vpa}(y,a,f) = offset(1 * \ln(I'_{(a,y,f)})) + \ln\left(\frac{1}{q_{a,f}}\right) \tag{3}$$

where «offset» means that the regression coefficient in front of the abundance index logarithm is equal to 1, while the standard error estimation of the inverse value of the catchability coefficient is estimated as the standard error of the regression parameter (a free term of the equation).

If the assumption of the constant variance (standard error) by years is invalid, application of the above equations will be incorrect. Therefore, in this case the method of weighted regression is to be applied instead of the traditional method, where the inverse values of standard errors of abundance indices, known from observations, can be used.

Two approaches can be used for direct estimation of standard errors of $N_{est}(y,a,f)$ values (or their variances), obtained by means of the weighted regression. The first approach presumes assessment of variance in the left part of equations (2) and (3) using variance or covariance matrix of components from the right part of these equations. The second approach presumes application of regression relationships between the logarithm of the mean abundance index and the logarithm of its standard deviation.

Apparently, the second approach is simpler and preferable, since it allows reducing the effect of sampling errors in estimation of abundance indices standard errors.

Definition of the weighted regression equations, where the accuracy estimates of the abundance indices are used as the weighing factors, is possible if the survey design has not been changed during a certain time period, i.e. the observation system remained stable. If this assumption is not valid, e.g. the number of hauls in strata was increased or decreased, which, naturally, affected the estimates accuracy, the direct estimates of standard errors shall be used for the entire time interval considered. At the same time, the variance of abundance estimates in the equations (2) and (3) can be calculated from the known "observed" variance values and from variances of parameters in these regression equations.

In our case the new version of XSA is presented as XSA with the weighted regression, which provides for application of regression relationships between the logarithm of the mean abundance index and the logarithm of its standard deviation. These regression relationships for herring and sprat were obtained from the data of BIAS 2004–2006 and became the basis of reproduction of abundance indices by age groups within the entire time interval of XSA models tuning, assuming a similar design of BIAS surveys during the whole time interval considered (ICES, 2002).

To investigate XSA method with the weighted regression the authors developed a new version of the software package implementing XSA method. The ICES software, used in the Baltic fish stocks assessment, the traditional version of XSA is applied based on the assumption of the constant variance of abundance indices by years and age groups (Darby, Flatman, 1994). The modification of ICES software consisted in replacement of the traditional linear regression to the regression with the known accuracy of predictors-abundance indices (Gasyukov, 2005).

The practical assessment of fish stocks by means of the new version of XSA method was fulfilled for herring of the Central basin (without the Gulf of Riga) and sprat of the Baltic Sea using the data applied by ICES WGBFAS in assessment of these species stocks in 2009 (ICES, 2009). These data included the standard quantity of files required for ICES software in XSA method implementation (Darby, Flatman, 1994): the total catch in tones, catch in specimens by age groups and fishing years, natural mortality rates, proportion of mature fish by years, mean fish weight in catches and in the stock. The data used in XSA tuning covered age groups 1–8 of sprat within the years from 1983 to 2008, and age groups 1–8 of herring within the years from 1982 to 2008. The oldest age group was considered as the plus-group. The basic options used in the new version of XSA were also similar to the options used by ICES WGBFAS-2009 (ICES, 2009).

The comparative analysis of the Baltic commercial fish stocks obtained on the basis of the new and traditional versions of XSA with the same input data are presented below.

Application of two XSA versions results in different estimates of herring and sprat stocks:

- *For sprat:* Estimates of recruitment by years are characterized with the high variability, while in some years (2004, 2006 and 2007) the traditional version of XSA gives the higher recruitment values as compared to the new version of the method. The similar conclusion may be made concerning the dynamics of the total and spawning biomasses. At the same time, the new version of XSA gives the higher estimates of fishing mortality. Comparison of this with recruitment and biomass estimates allows concluding that the traditional XSA method provides for the more optimistic estimate of sprat stocks state for the recent years.
- *For herring*: The comparison of two assessment versions results even in more explicit conclusions: the estimates of all parameters (recruitment, total and spawning biomass) based on the traditional XSA version exceed the respective estimates obtained with the weighted XSA version. In some years (2004, 2005 and 2006) the spawning biomass estimates exceeded more than twice the respective values obtained with the new XSA version. At the same time, the mean fishing mortality rates estimated with the traditional XSA version were above two times lower than the respective values obtained with the new XSA version. Like in the case with sprat, the conclusion can be made that the estimates for herring obtained by WGBFAS are more optimistic than the actual ones.

The empiric basis of relationship between the standard error of abundance indices and abundance indices value is the important argument in support of these conclusions.

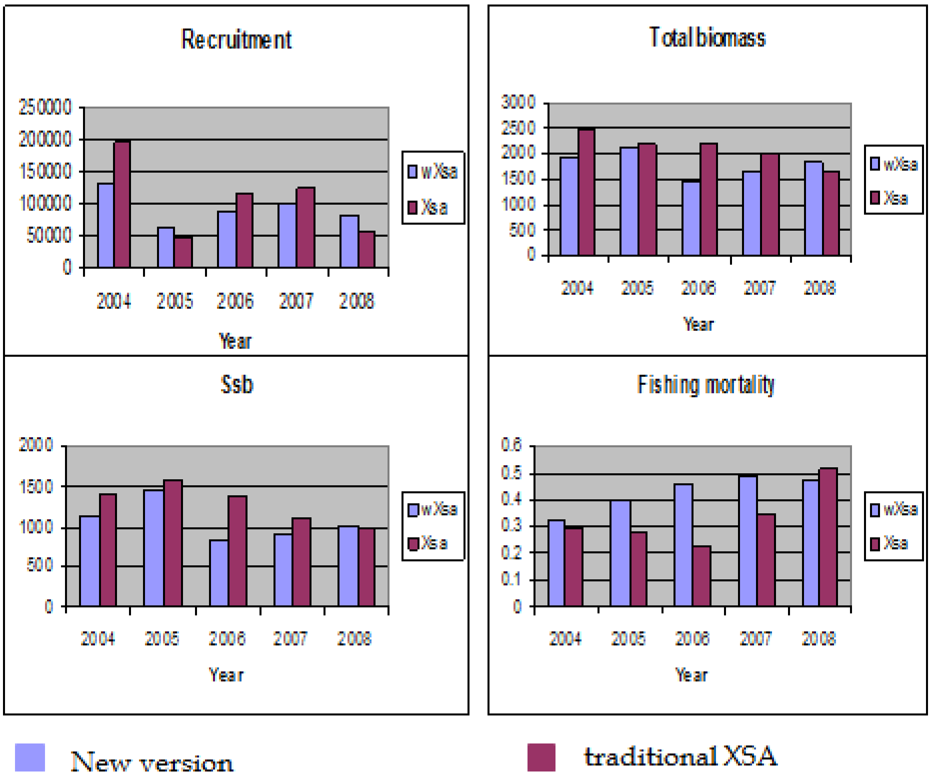


Figure 3. Sprat stock and population parameters in the Baltic Sea calculated by traditional XSA and XSA with weighted regression (wXsa).

Figure 4. Herring stock and population parameters in the Baltic Sea calculated by traditional XSA and XSA with weighted regression (wXsa).

Figures 5 and 6 show the retrospective analysis for sprat stock estimates. It is possible to state that the estimates obtained by XSA with weighted regression produce lesser errors in retrospective bias then traditional XSA. Therefore the reasons and possible explanations for such type of errors can be found in some misspecification of the stock assessment model.

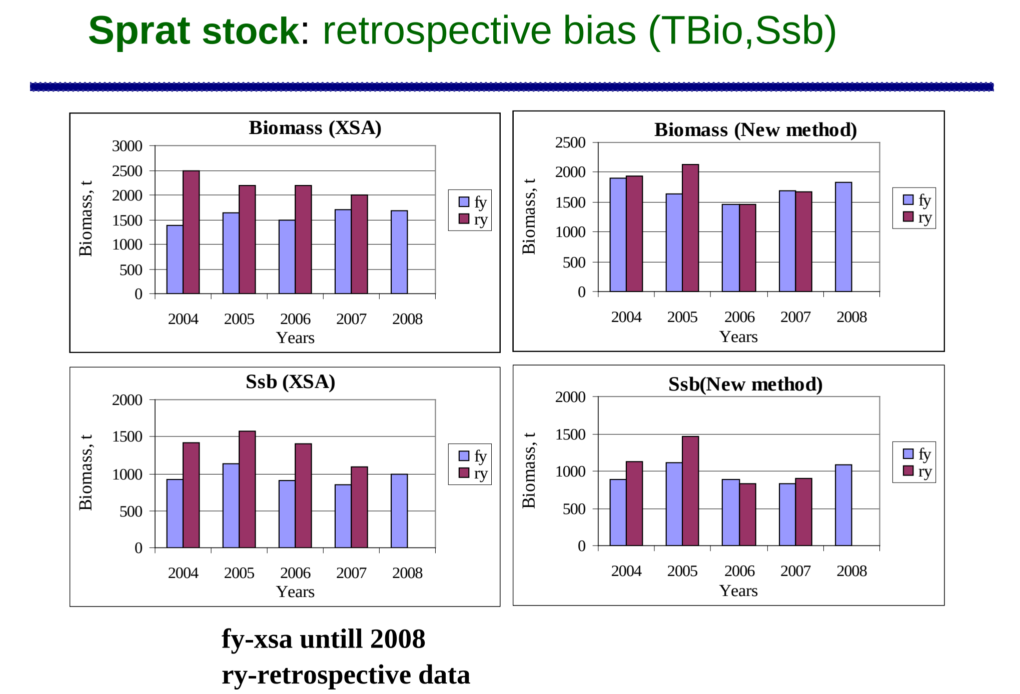


Figure 5. Sprat retrospective estimates of total biomass and spawning biomass obtained by traditional XSA and XSA with weighted regression.

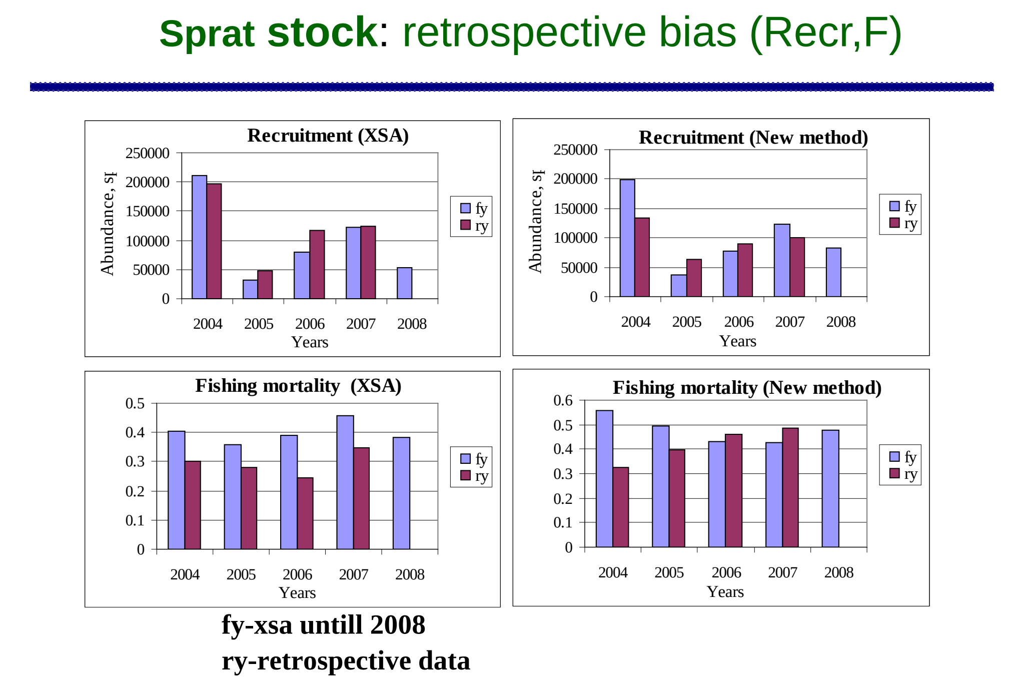


Figure 6. Sprat retrospective estimates of recruitment and fishing mortality obtained by traditional XSA and XSA with weighted regression.

CONCLUSION

Application of the new XSA version considering the variability of abundance indices variance by years may result not only in new estimates of fish stock and population parameters (recruitment, total and spawning biomasses, mean fishing mortality rate), but also may change the temporal trends and retrospective bias in fish stock dynamics.

Estimating uncertainty in abundance indices based on acoustic surveys and subsequent integration of these estimates into the stock assessment models are very urgent in view of ICES initiatives to revise stocks assessment methods in compliance with the Strategy Research Plan .

The software used by the ICES working groups for stock assessment should contain the options which give the possibility to take into account the indices variance variability in time

Acoustically derived indices needed for stock assessment purpose should be accompanied with uncertainty estimates. **How can be it included in the real practice of acoustic surveys?**

Possible recommendations from this work are:

1) Scientific surveys (trawl and acoustic) of the fishery stocks should be accompanied by the statistical estimates of abundance indices (at least variance).
2) The software used by the ICES working groups for stock assessment should contain the options which give the possibility to take into account the variance variability in time
3) In our opinion estimating uncertainty in abundance indices based on the Baltic International surveys and subsequent integration of these estimates into the stock assessment models are very urgent in view of ICES initiatives to revise stocks assessment methods in compliance with the Strategy Research Plan.

REFERENCES

Efron, B. Non-traditional methods of the multivariate statistical analysis. M.: Finances and Statistics, 1986. 263 p.

Efron, B., Tibshiran,i R. An introduction to the bootstrap. New York.: Chapman & Hall. 1993. 436 p.

Darby, C., Flatman, S. Virtual Population Analysis. Version 3.1. User Guide. Copenhagen, Denmark. ICES press, 1994. 85 p.

Gasyukov, P. G. On the use of “shrinkage procedure” in the ICES program realizing the “Extended survivors analysis” (XSA) method // ICES Document CM 2005/ACFM:19 Ref. H. 2006. Pp.517.

ICES. Report of Baltic International Fish Survey Working Group (WGBIFS). Annex 2. Manual for Baltic International Acoustic Survey (version 072. 07.04.2000) // ICES Document CM 2002/G:05. 2002. Pp. 90–115.

ICES Science Plan (2009–2013). Copenhagen. Denmark. ICES press, 2008. 15 p.

ICES Report of the Baltic Fisheries Assessment Working Group (WGBFAS)// Document ICES CM 2008/ACOM: 06.2008.250p.

ICES Report of the Baltic Fisheries Assessment Working Group (WGBFAS)// Document ICES CM 2009/ACOM: 07. 2009. 230p.

Kasatkina, S. M., Gasyukov, P. G. Estimating uncertainty in the Baltic acoustic survey results applying geostatistics techniques and simulation // ICES Document CM 2006/I: 14. 2006. 17p

Kasatkina, S. M., Gasyukov, P. G. Quality of abundance indices based on international acoustic surveys in context of input data for stock-assessment models: example of Baltic International Acoustic Surveys // ICES Document CM 2009/N:12. 2009. 23 p.

Shepherd, J. G. Extended survivors analysis: An improved method for the analysis of catch-at-age data and abundance indices / J.G. Shepherd// *ICES Journal of Marine Science*.1999.V. 56. Pp.584–591.

Working Document 4

Preliminary assessment of white anglerfish southern stock using Stock Synthesis (SS3)

P. Sampedro[1] and C. Fernández[2]

[1]IEO, Centro Oceanográfico de A Coruña, Pº Mº Alcalde Francisco Vázquez Nº 10, 15001 A Coruña, Spain
[2]IEO, Centro Oceanográfico de Vigo, Cabo Estay-Canido, 36200 Vigo, Spain

Abstract

A first attempt of assessment of white anglerfish southern stock using Stock Synthesis (SS3) is presented in order to evaluate its potential use as an alternative assessment model to the current surplus production model (ASPIC). Model structure, input data and provisional model settings are described in the work. Although more effort is required for tuning the model, the fit and the preliminary results seem to indicate that the Stock Synthesis can be an appropriate model to assess this stock.

Introduction

Over the last four years, the assessment of white anglerfish southern stock has been conducted using a surplus production model. This production model has some advantages as it is relatively simple, there is few parameters to be estimated and thus is relatively quick to run. However its ability to capture particular details of the fishery under investigation is limited. The simple structural of the model does not allow for observational data such as length and/or age compositions to be employed. Consequently, potentially important information regarding individual gear selectivity and basic biological information is not able to be integrated directly into the model.

In this document a new model platform was explored to perform the assessment of white anglerfish southern stock. The Stock Synthesis (Methot, 2000) is an integrated assessment model that is able to handle large amounts and different kinds of data and is flexible with respect to the underlying population dynamics and to the number of parameters that can be estimated. Stock Synthesis has been widely used and tested for stock assessments, especially in the US west coast. The use of the SS3 (Methot, 2011) for this stock is going to be proposed in the ICES benchmark scheduled in 2012. As a previous step, and for its review by the working group, a preliminary assessment using this model is presented.

Material and Methods

The following model structure, input data and settings were used in the analysis and population dynamics calculated from 1989 to 2009. The quarterly based data (landings, LPUE and length–frequency) was used in the SS3 calculation:

Model structure

- 1 sex (both sexes combined)
- 1 area
- 4 seasons per year (model time-step is quarter)

- 1 growth pattern

Model input data

The assessment model includes four 'fleets' defined on the basis of gear type and area:

- SPTR8C9A: Spanish Otter BottomTrawl in ICES Subdivisions 8c and 9a
- SPART8C: Spanish Artisanal in ICES Subdivision 8c
- PTTR9A: Portugal Trawl in ICES Subdivision 9a
- PTART9A: Portugal Artisanal in ICES Subdivision 9a

Landings in weight (Figure 1) and length frequency distributions were the inputs.

Three abundance indices were considered:

- SPCORUTR8C: Spanish Otter BottomTrawl in Subdivisions 8c and 9a
- SPCEDGN8C: Spanish Artisanal in Subdivision 8c
- SP-SGF: Spanish Ground Fish Survey

Model settings

Stock–recruitment relationship. Recruitment was modelled assuming a Beverton and Holt curve and (h, R0) was defined as the parameters instead of (a, b) in the B-H function. Steepness value (h) was fixed to 0.999 and estimated R0. Annual recruitment was distributed by quarter, being first and second quarters the most important seasons in recruitment.

Growth curve. The von Bertalanffy growth curve was used. The provisional parameters were Linf=140.7 cm (estimated) and K=0.11 (fixed; Figure 2).

The weight at length relationship was: $W=0.0000270*L^{2.8390}$ (BIOSDEF 1998)

The natural mortality is assumed to be age and time independent and equal to 0.20 yr^{-1}.

Maturity of white anglerfish is assumed to be logistic in shape and a function of length. The parameters, externally estimated, were as follows: length at 50% maturity = 55.4 cm and a slope of linearized logistic equation = -0.12; Figure 3).

Selectivity patterns (relative exploitation patterns). Selectivity is assumed to be length-based for all fleets. It was adopted a dome-shaped double-normal selectivity with six parameters in all commercial fleets and abundance indices, and estimated unknown parameters in each case. Initial values were set up based on the actual size frequency in each fishery (Figure 4).

Results

The estimated retention curves for the commercial fleets are shown in Figure 4. The SS3 estimated high selectivity for the larger length classes for the trawl fisheries from Portugal (PTTR9A) probably due to the observation of a very small number of larger fish in the landings composition. On the other hand, for Spanish trawl and artisanal fisheries the relative selectivity in larger sizes drops to zero. The shape of these curves cannot appear totally reasonable and more effort will be needed to tune fishery retention patterns.

The model was able to capture the general trend and the interannual variation for the survey abundance index and the commercial fleet SPCORUTR8C (Figure 5). The model only follows the overall trend in the commercial abundance index SPCEDGN8C, the high observed values in 2003, 2005 and 2007 have not been captured by the model in this fleet.

The SS3 model estimated that SSB has steadily increased since 1996 to about 12 000 mt (Figure 6). A very low recruitment period, during five consecutive years (1995–1999), is detected in the estimated recruitment time-series. Since 2000, alternating years of good recruitments (2001, 2004 and 2009) and low recruitments (2003, 2005, 2007 and 2008) are observed. The pattern of Fbar (40–85 cm) shows fishing mortality decreasing to its lowest level on 2001 and, after a small recovery period, Fbar presents a declining trend since 2005.

A comparison of the biomass trajectories from the SS3 model (SSB) and ASPIC results (total biomass) revealed different trends on biomass. Since 1996, SS3 biomass estimates are quite higher than the production model estimates, reaching the maximum value of its time-series in 2009. The shorter time-series used in SS3 model (1989-2009), where the beginning period of the stock fishery is missed, could have an effect on the estimates obtained.

Conclusions

The analysis of the retention curves indicated that more effort in this subject is necessary to tune the SS3 model. In order to obtained reliable results, the input data time-series should be update to incorporate information from early years. Although the present SS3 model for white anglerfish needs to be improved, its results and diagnostics seem to indicate that this can be an appropriate model to assess this stock.

References

BIOSDEF. 1998. Biological studies of demersal fish. Ref.: EU, DG XIV, Study Contract 95/038.

ICES. 2010. Report of the Working Group on the Assessment of Southern Shelf Stocks of Hake, Monk and Megrim. ICES CM 2010/ACOM:11, 607 pp.

Methot, R. D. 2000. Technical description of the Stock Synthesis assessment program. National Marine Fisheries Service, Seattle, 46 pp.

Methot, R. D. 2011. User manual for Stock Synthesis: Model Version 3.20 (Updated January 21, 2011), 165 pp.

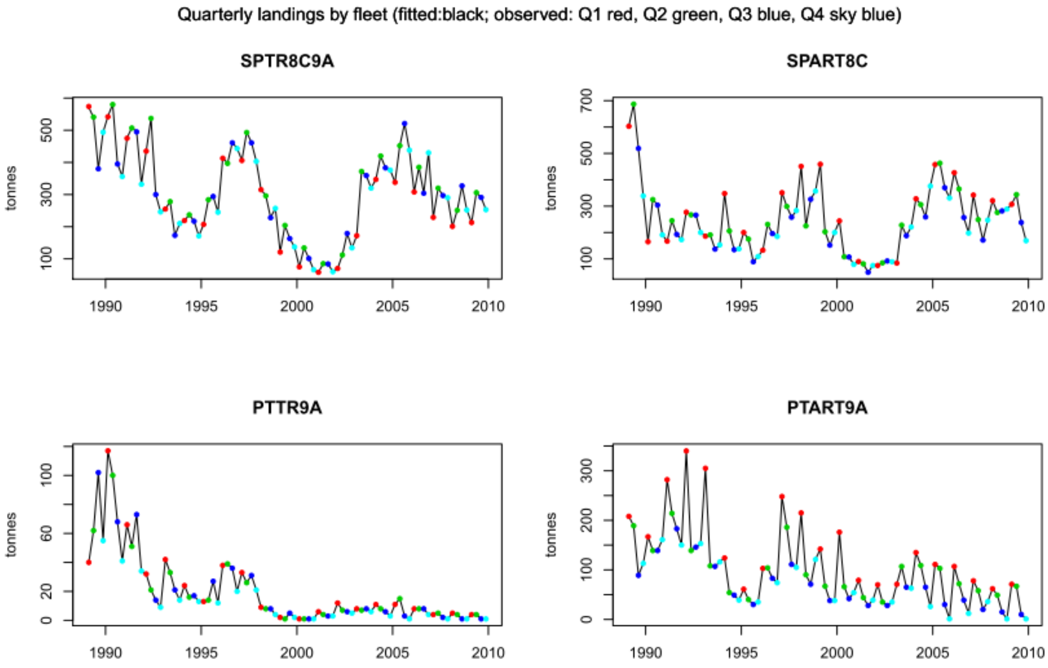


Figure 1. Quarterly landings for the four commercial fleets used in SS3 assessment.

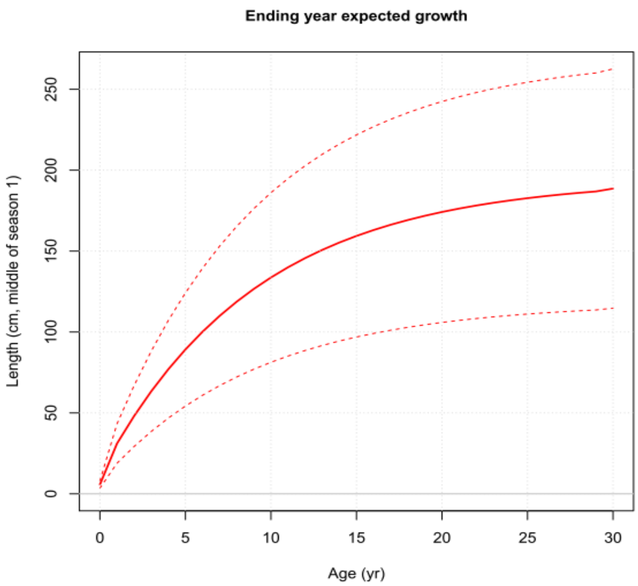


Figure 2. Growth curve with ~95% interval (dashed lines) indicating the expectation and individual variability of length-at-age.

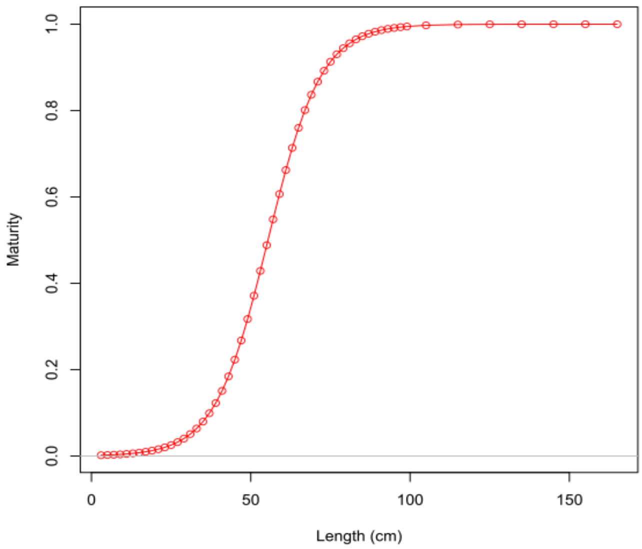


Figure 3. Maturity ogive based on length fixed in SS3 assessment.

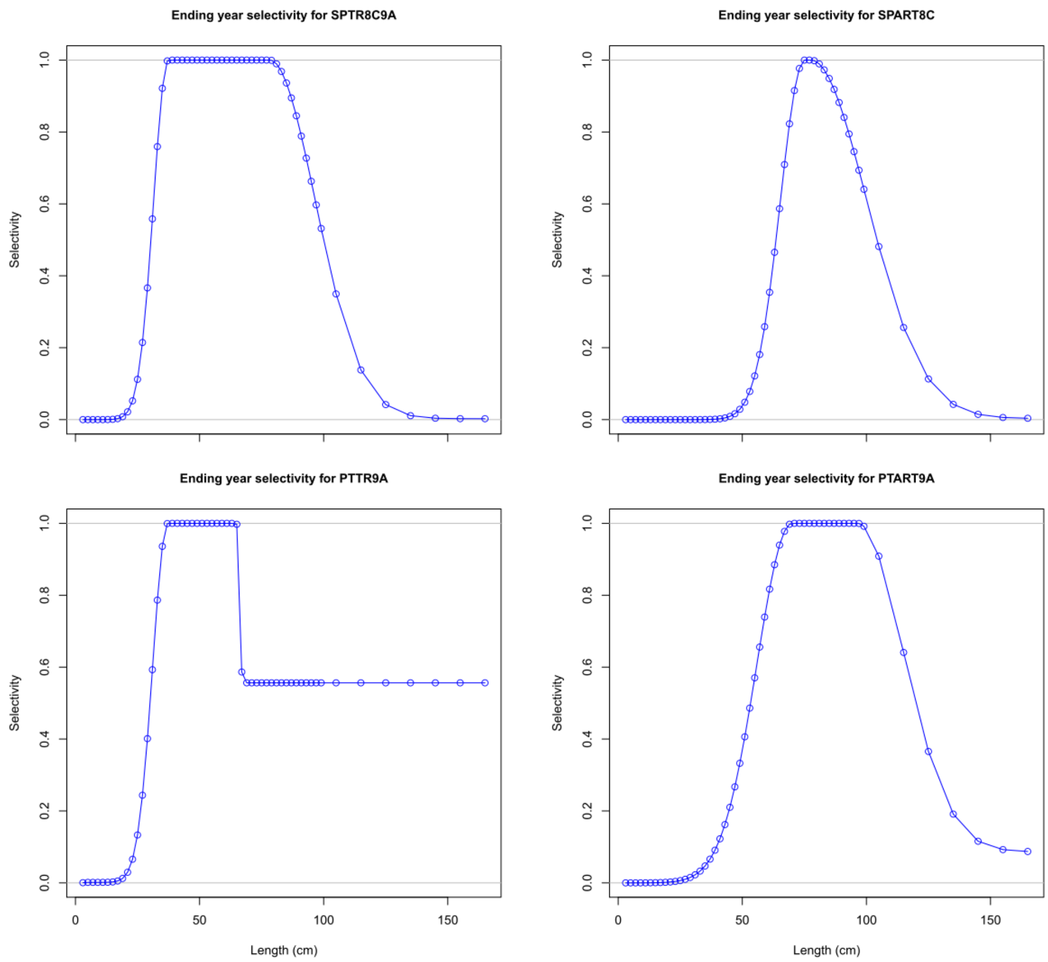


Figure 4. Estimated size-specific relative exploitation pattern for the four commercial fleets.

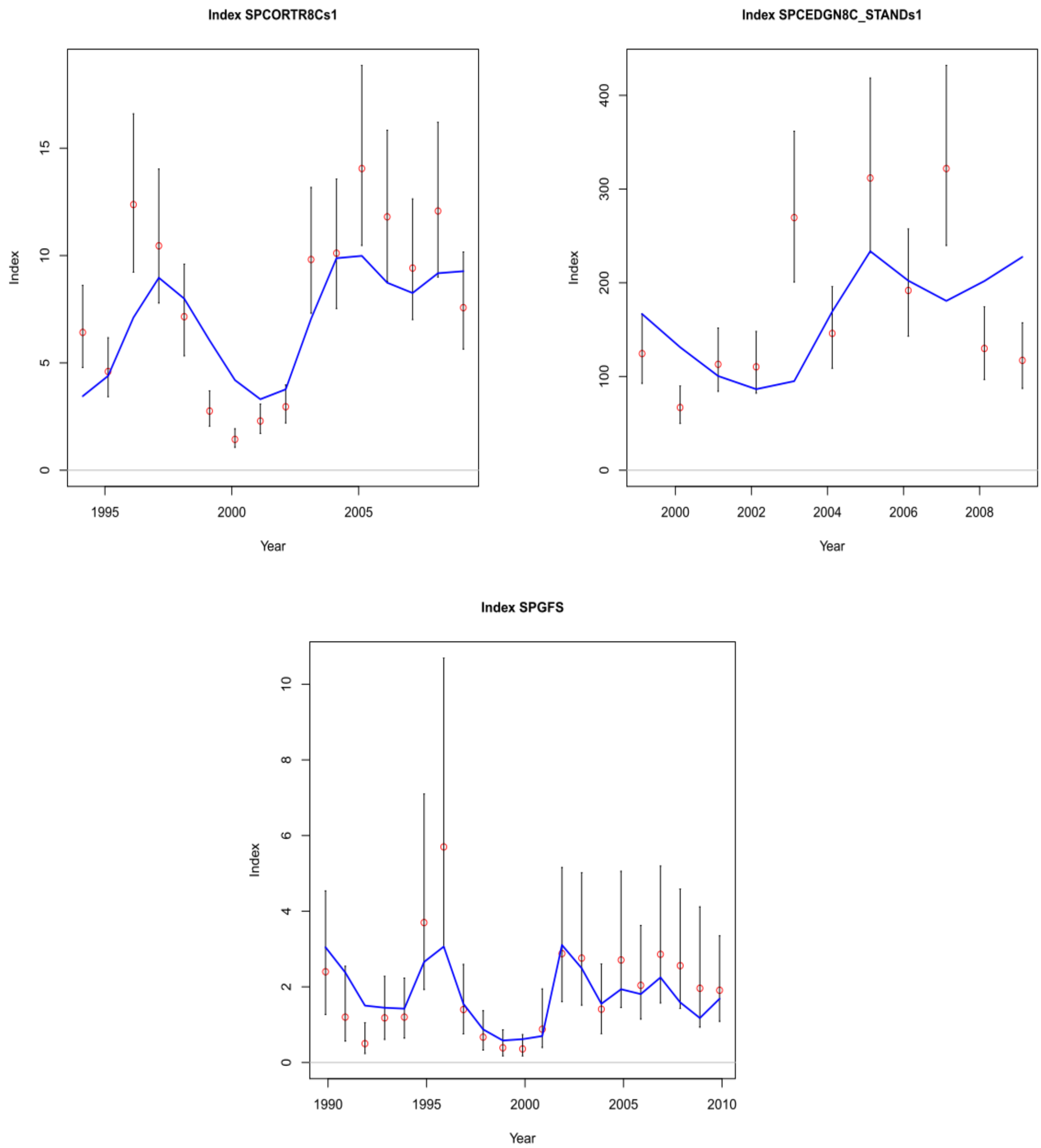


Figure 5. Fits to the two commercial fleet (top) and survey (bottom) abundance indices. Predicted (line) and observed (circles) indices.

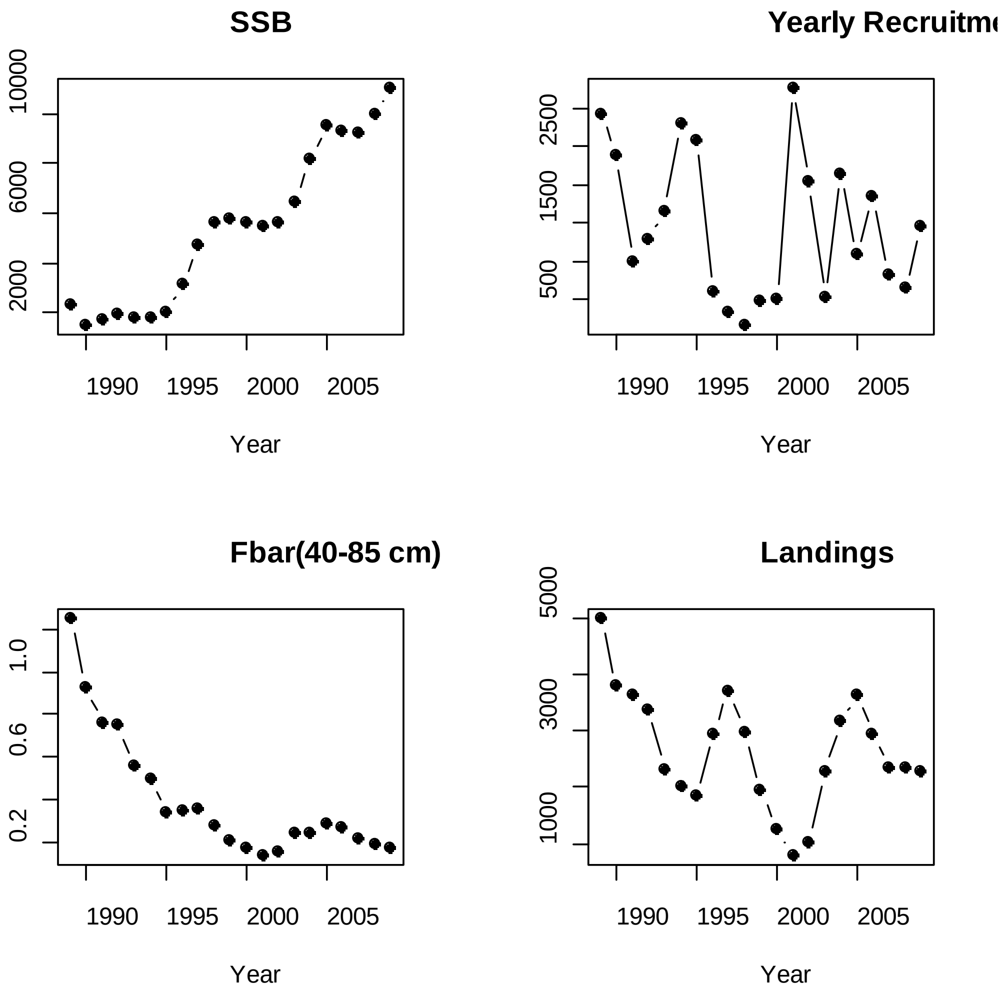


Figure 6. Stock trends derived from SS3 assessment results.

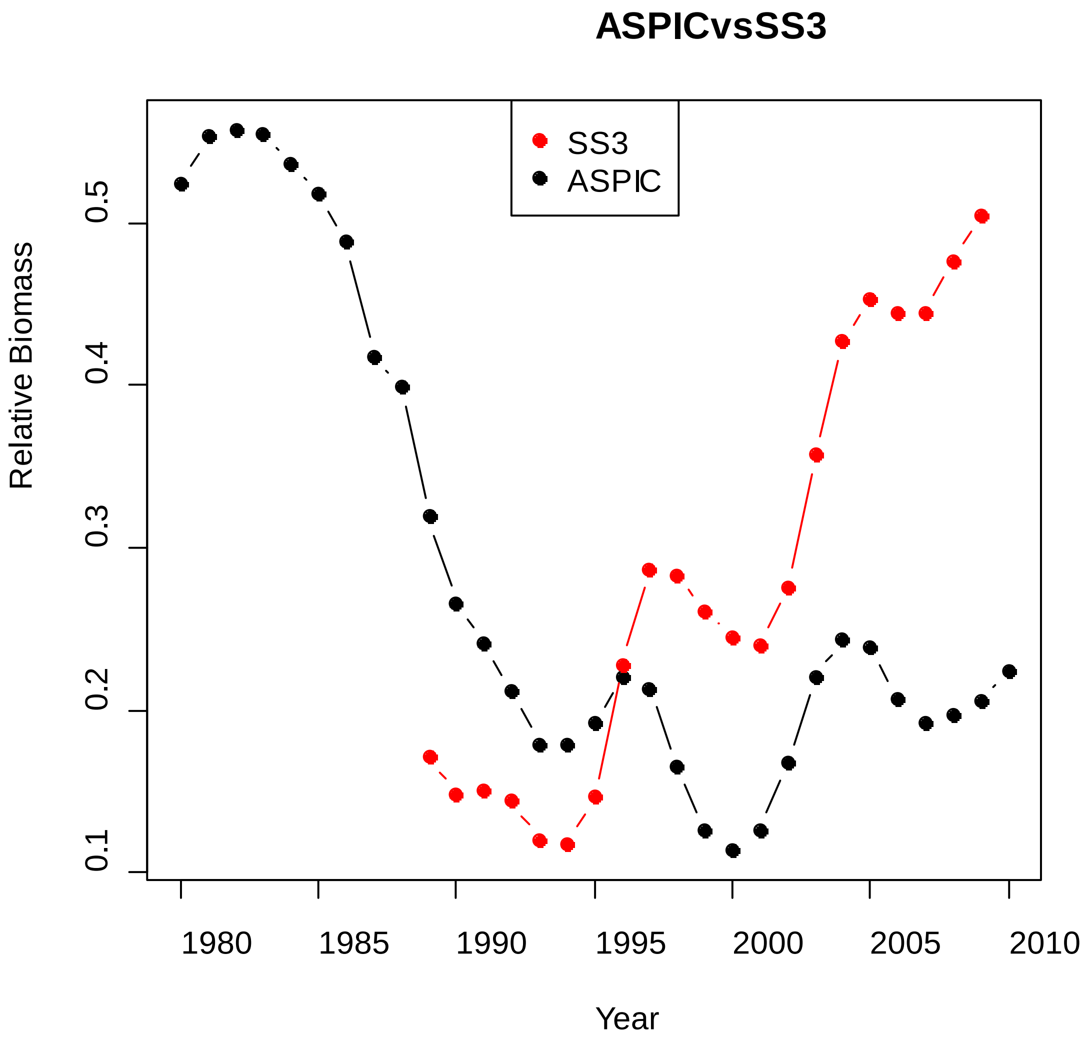


Figure 7. Comparison of biomass trend from ASPIC (ICES, 2010) and SS3 assessments.

Working Document 5

Applying a Bayesian model incorporating discards in the assessment of four-spot megrim (*Lepidorhombus boscii*) Southern stock

Esther Abad, Carmen Fernández, Nélida Pérez.

Centro Oceanográfico de Vigo. Instituto Español de Oceanografía (IEO). Apdo. 1552. 36200 Vigo (Spain).

Since 2003 when the DCF started at European level by countries, discards data are available for many stocks. Four-spot megrim is traditionally assessed with XSA (extended survivor analysis) which does not include discards. For this species, discards in number are very important, being around the 60% of total catch. A Bayesian model incorporating discards was realized for the hake stock in ICES Divisions VIIIc and IXa by Fernández *et al*. (2010). This model was also designed to produce a complete time-series of discard estimates. Final run of the model is compared with results from XSA performed in the working group of 2010, showing that mayor differences are in the fishing mortality for younger ages, being higher incorporating discards data.

MATERIAL AND METHODS

Data are the same used in the working group for the assessment of the Southern stock of *L. boscii*.

Landings data are provided by National Government and research institutions of Spain and Portugal with data since 1986. Age compositions of landings are based on annual Spanish ALKs. Since there is no landing for age 0, landed numbers-at-age are presented from age 1 to 7+.

Discards estimates are available for Spanish since 2003 and before this year, there are data only for 1994, 1997, 1999 and 2000. Discard numbers-at-age are presented from age 0 to 5. Portuguese discards are assumed to be zero at this moment, but they will be able to be incorporated if they are available.

To tune the model two indices are available, one commercial fleet and one survey index. A Coruña trawl fleet (SP-CORUTR8c) contributing with data of effort and LPUE till 1999 due to changes in the fishing gears and the Spanish groundfish survey (SP-GFS), available since 1983, are the two indices. Numbers-at-age are from 3 to 6 in the trawl fleet and from 0 to 6 in the research survey.

<u>Assessment model</u>

Model is described in Fernández *et al.* (2010). It is a Bayesian model computed in the free software WinBUGS for simulating the posterior distribution via Markov chain Monte Carlo (MCMC). The population dynamics is based on the usual equations for closed population and the rate of fishing mortality is disjointed in two terms, one related to landings and other related to discards.

$$F(y,a)=f(y).(s_L(y,a)+s_D(y,a))$$

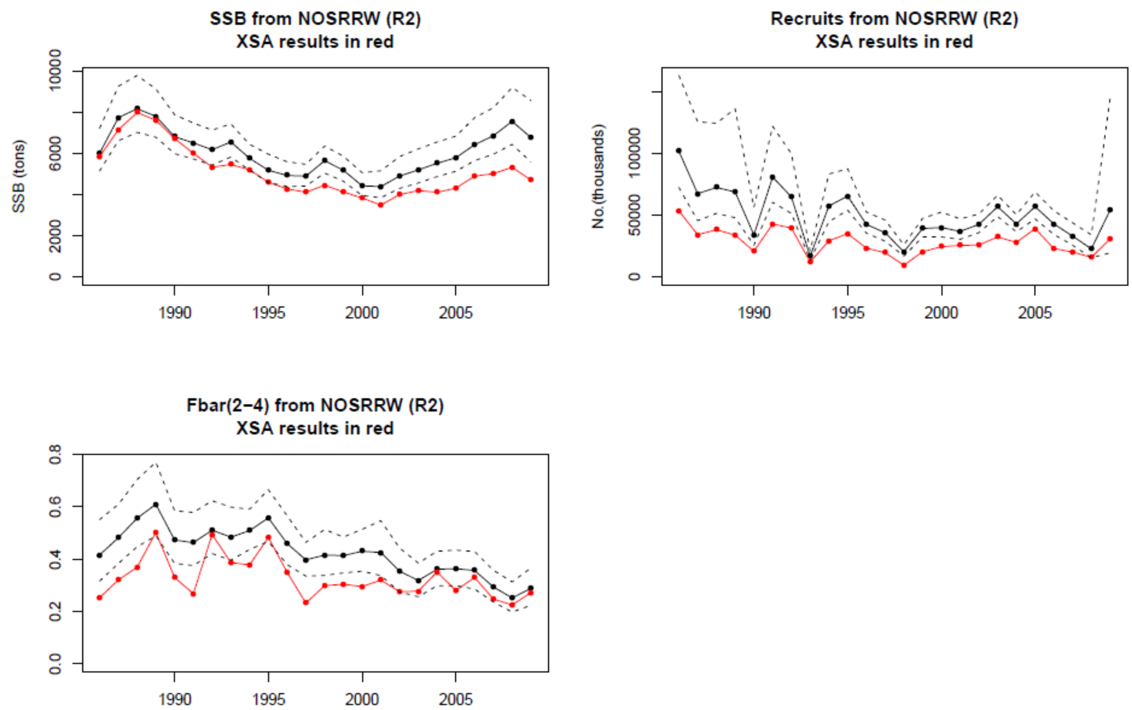
Fishing mortality is the result of the product of fishing effort and the exploitation pattern, being the last one age dependant. To obtain landed numbers-at-age, the model applies the Baranov catch equation. The two series correspond to the commercial fleet and the research survey, are used to obtain relative indices of abundance-at-age. All formulations are showed in the paper mentioned above.

The unknown parameters of the model are assigned prior distributions. These distributions are set with two values, the median and the precision. As better is our previous parameter knowledge, setting values have lesser variability.

The number of iterations used to fit the model was the same as in the hake model. The fitting was made using MCMC to simulate the posterior distribution with 112000 iterations. The first 32000 were not achieved and 5000 iterations from the rest were kept.

RESULTS AND DISCUSSION

Figure 1 shows results of SSB, Recruits and Fbar from the model comparing with those obtained with XSA in the last assessment. Trends are very similar. In SSB, first years are more coincident than last years', where Bayesian model estimates greater values of SSB. Including discards, looks to have an effect on the recruitments estimation, resulting in higher values almost all the time-series. Discards contribute with important amounts of earlier ages. Fishing mortalities also present the same trend, but last years are more similar than the beginning of the series.



In figure 2, the evolution of fishery age selectivity is presented comparing with the same results from XSA. As it is expected, there are more differences in the two first ages when discards are introduced in the model. First of all we have fishing mortalities for age 0, which did not appear in the XSA results because there are not commercial landings for this age. For age 1, values are now higher than in the XSA model and a little higher in age 2. In the rest of ages, values are more or less coincident.

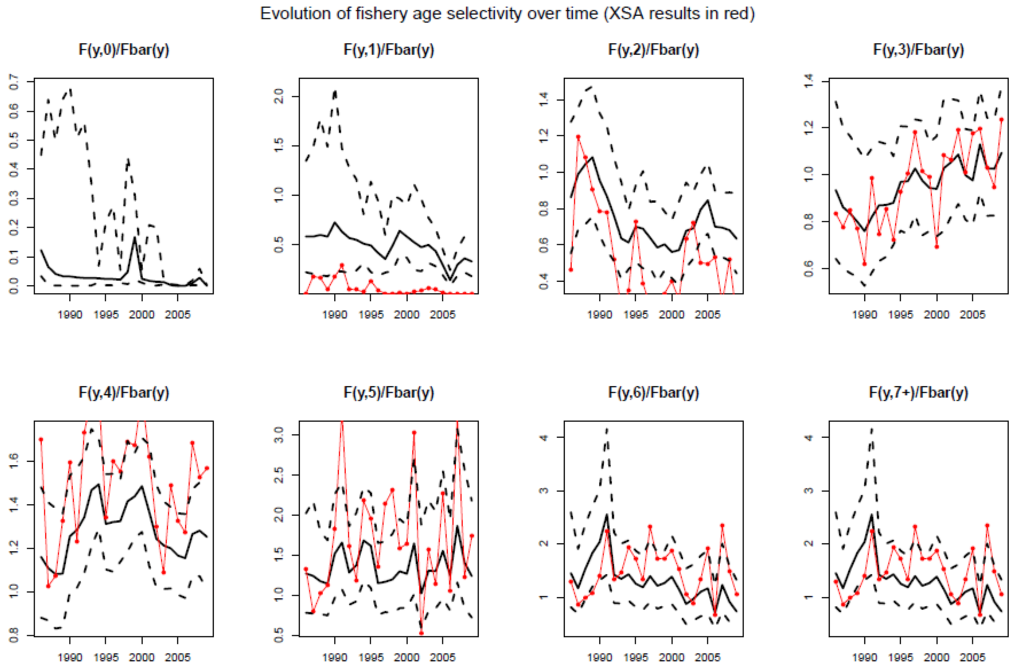


Figure 3 shows the same as figure 2 but with all ages together. Red line is corresponding to the last assessment year (2009), where the exploitation pattern indicates the biggest fishing pressures on ages 3 to 5.

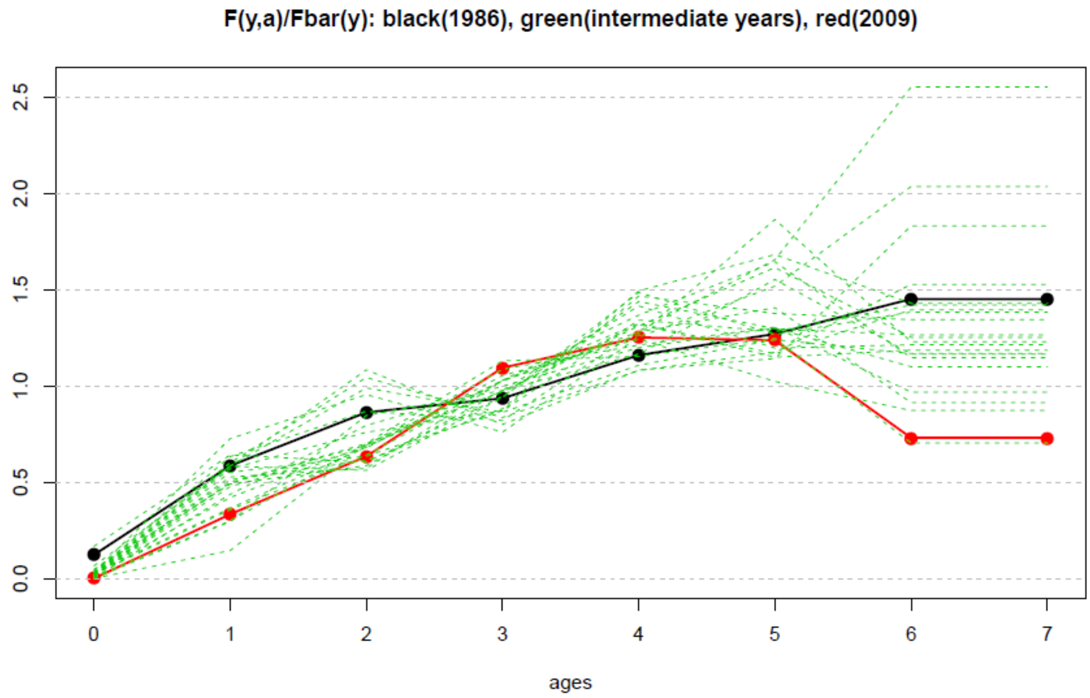
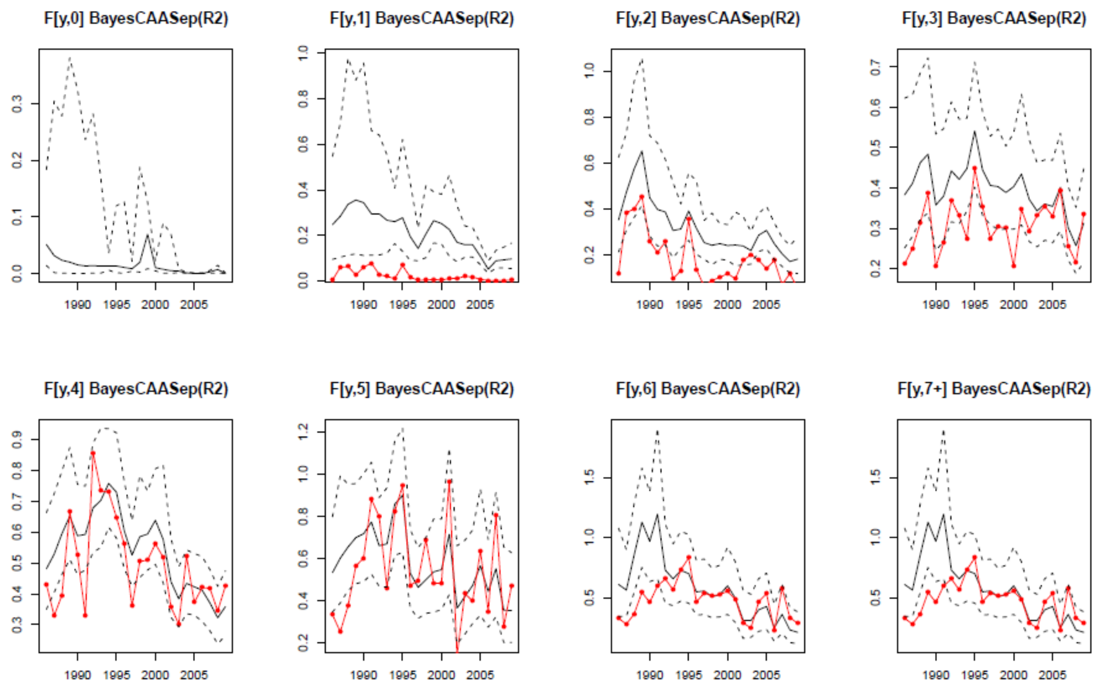


Figure 4 shows the values of fishing mortalities for all years and ages. As in the case of de fishery age selectivity, the biggest changes are for age 0 and 1, being a bit higher of XSA results for ages2 and 3 and more similar for the rest of the ages.



In figure 5 the probability that the fish being caught are discarded for the different years and ages is presented. There is no evidence of the effect of the progressive enforcement of the MLS for this stock since 2000. Apparently, there is a decrease of the probability after this year for ages 2–4, but a high increase is detected following it for ages 2 and 3.

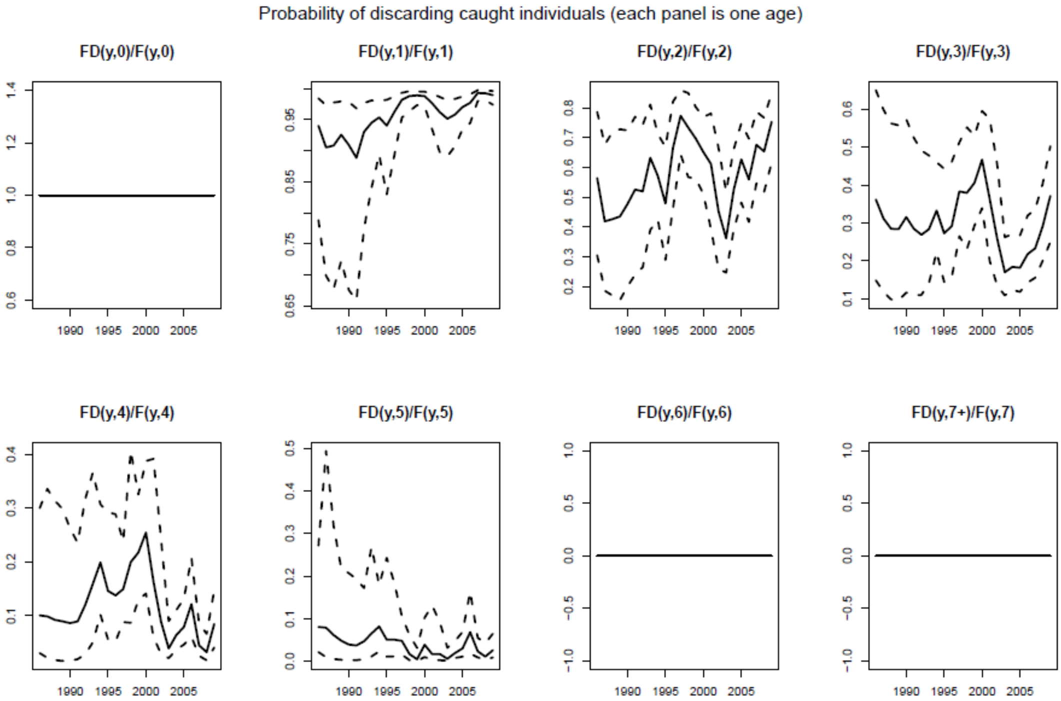
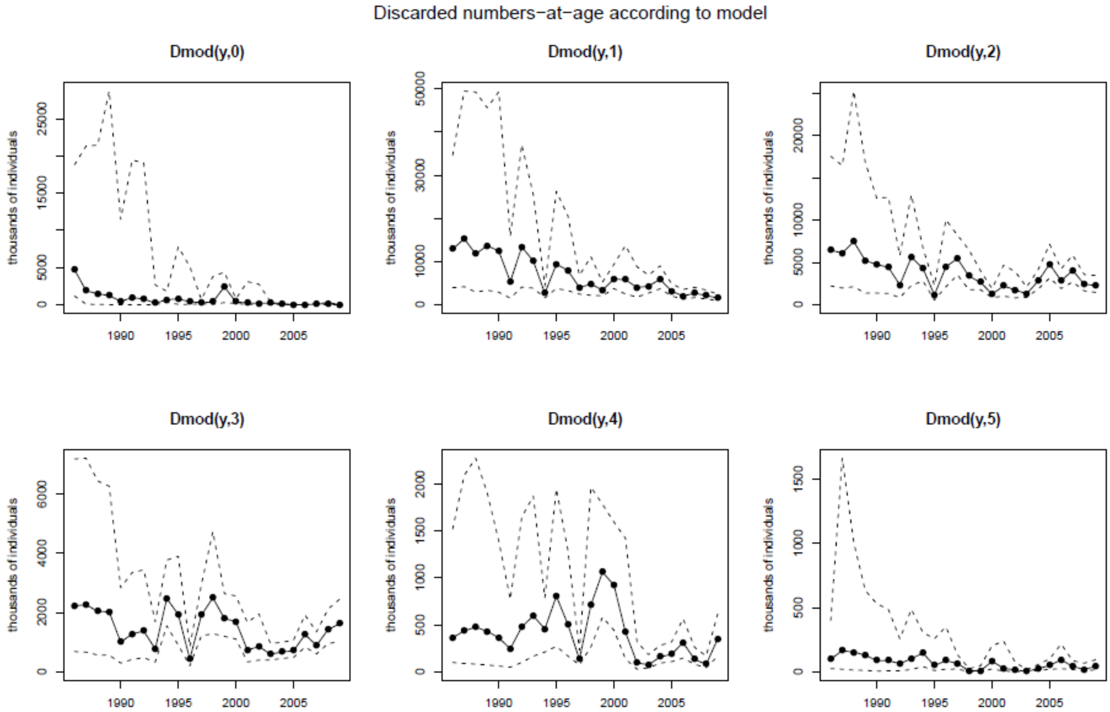


Figure 6 presents discarded number-at-age as result of the model. Very high variability can be observed in the first years of the time-series, when only a few years have observed data. Since 2003 model fits better because all years have been sampled.



From Figures 7 to 11, standardized residuals are presented. There is nothing relevant to discuss about this values, and its evolution during the time-series can be observed in next figures.

Figure 7:

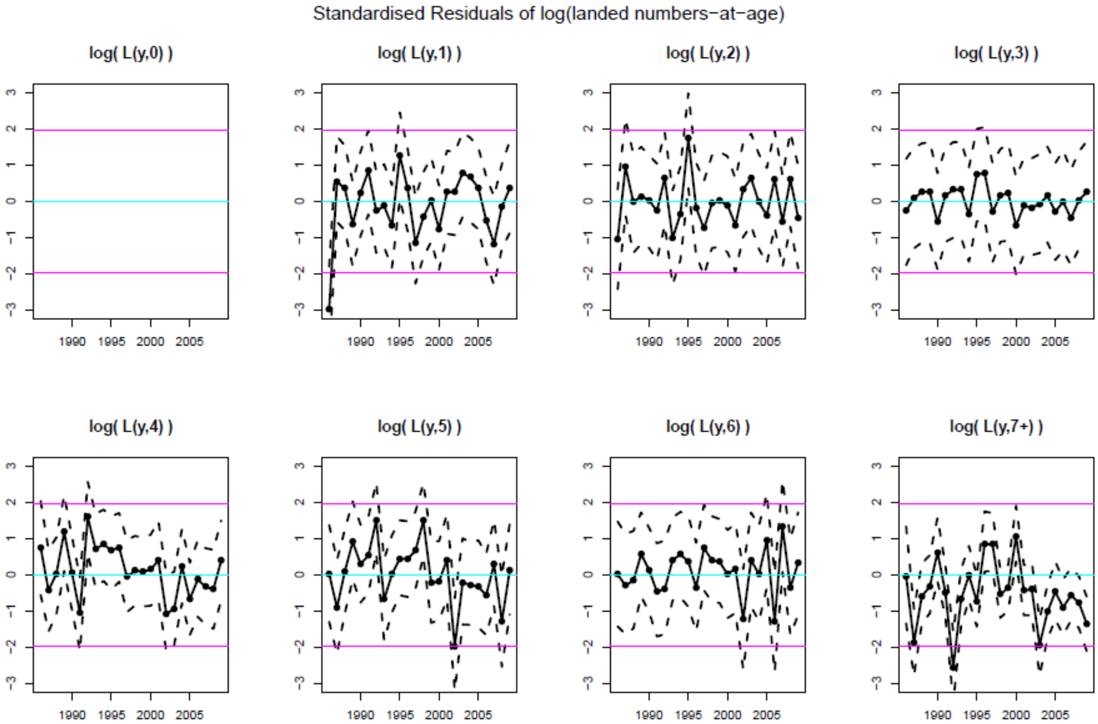


Figure 8:

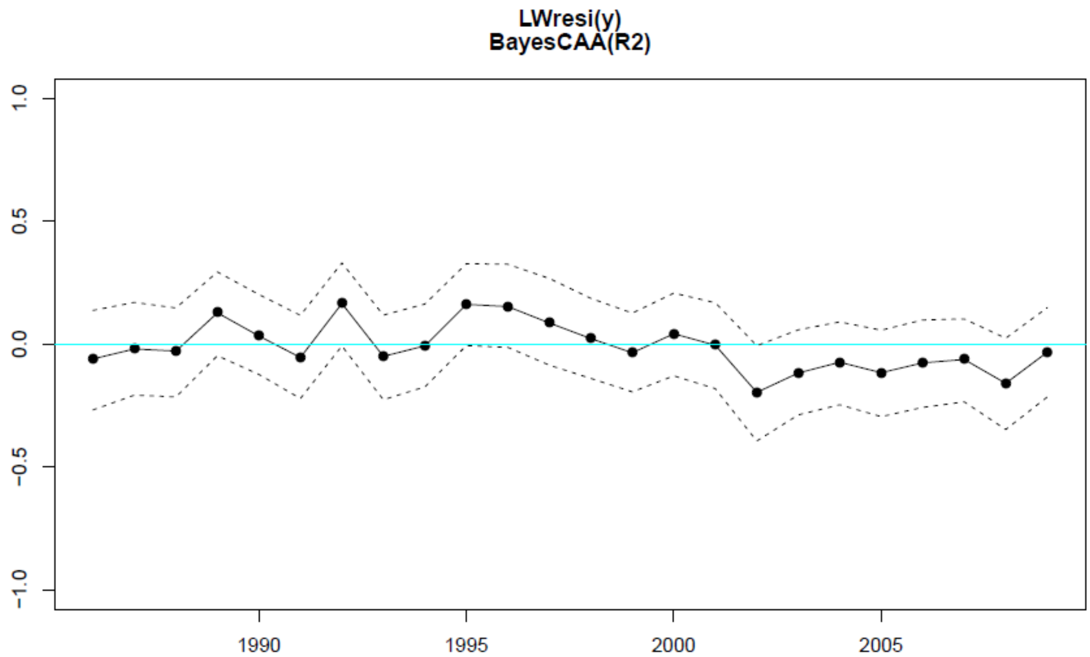


Figure 9:

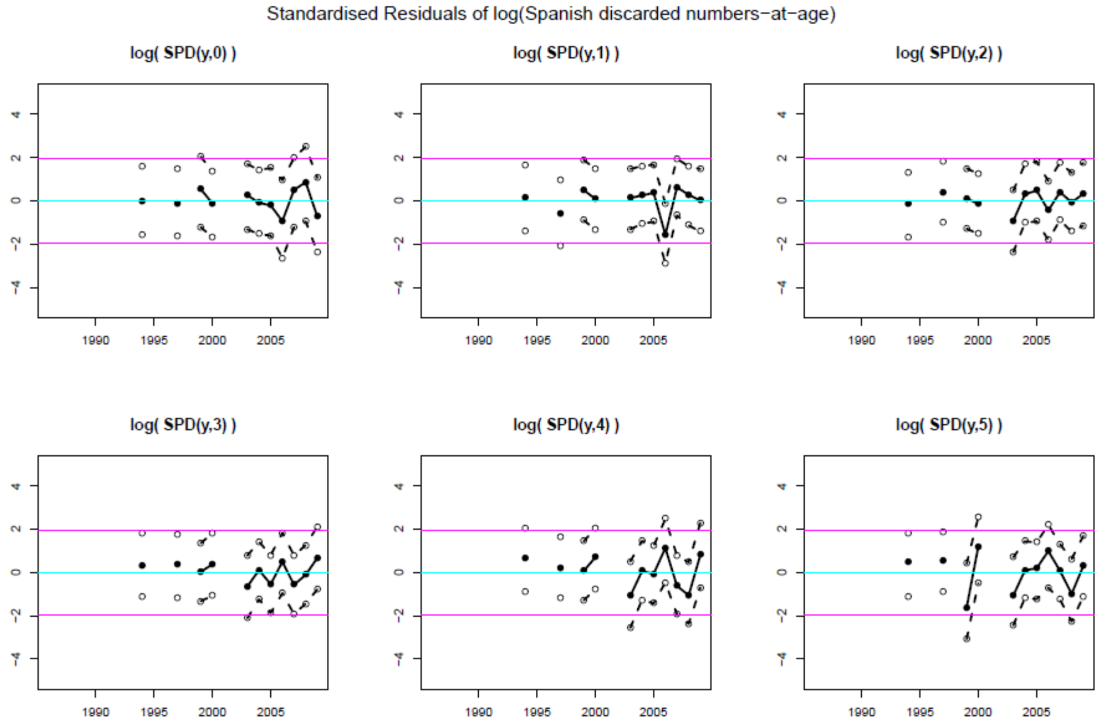


Figure 10

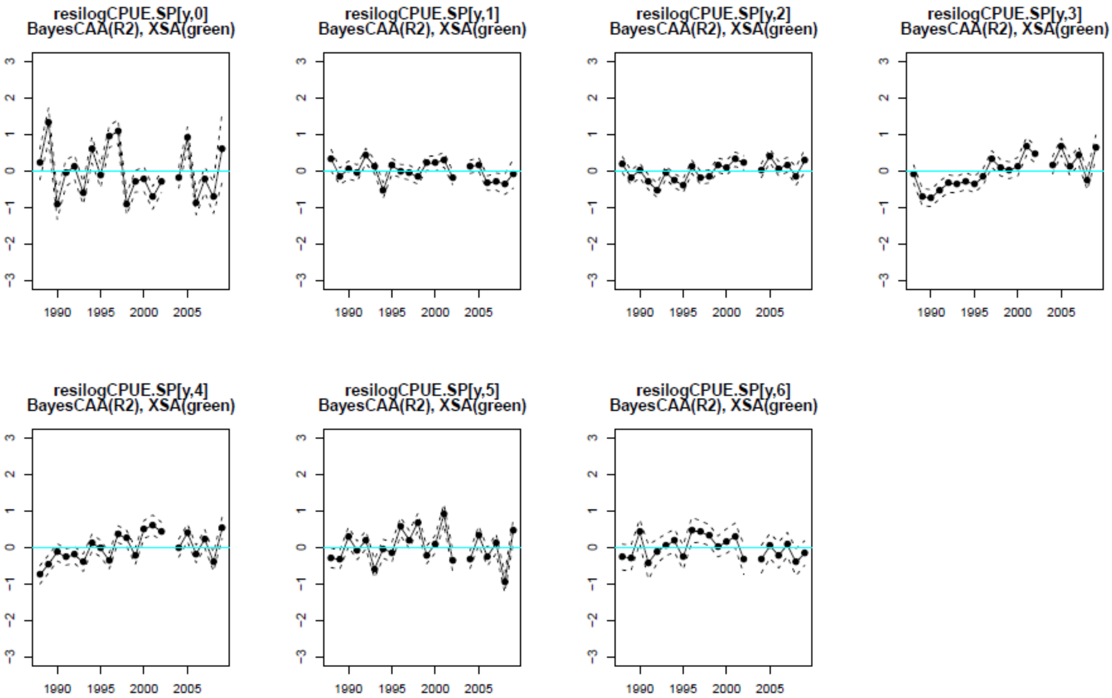
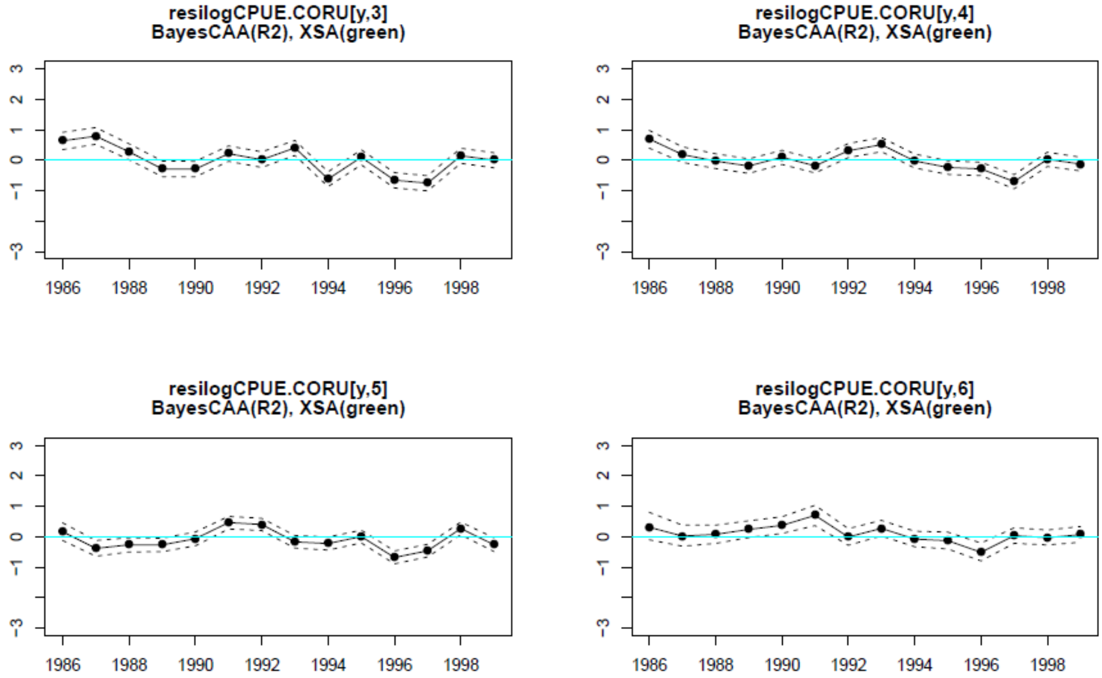
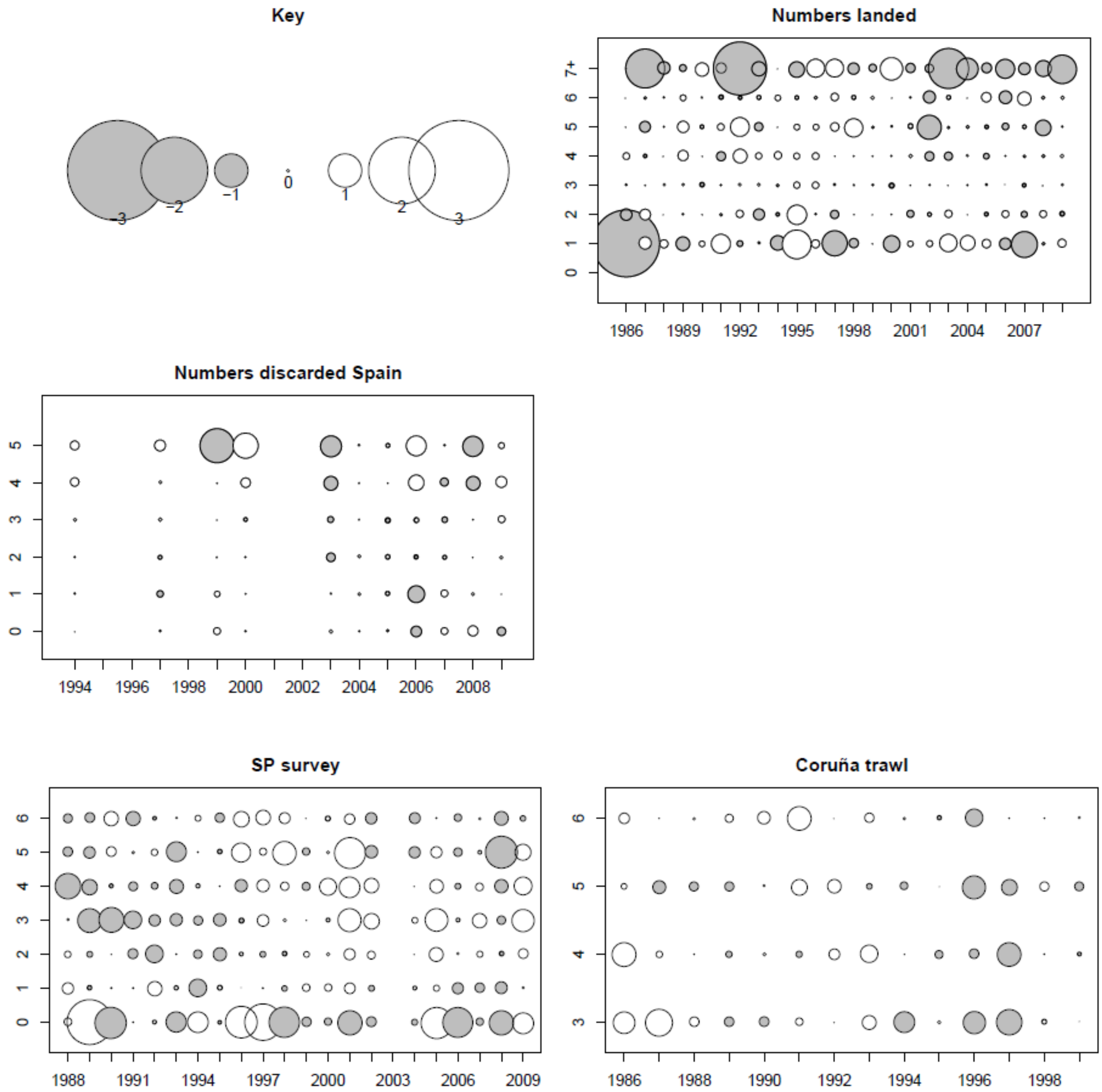


Figure 11

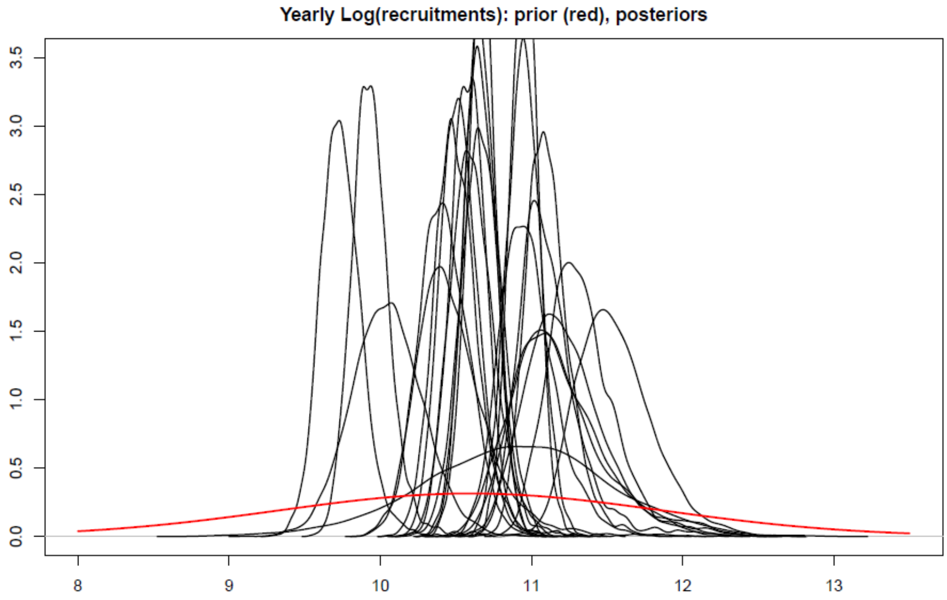


resilogCPUE.CORU[y,3]
BayesCAA(R2), XSA(green)
resilogCPUE.CORU[y,4]
BayesCAA(R2), XSA(green)
resilogCPUE.CORU[y,5]
BayesCAA(R2), XSA(green)
resilogCPUE.CORU[y,6]
BayesCAA(R2), XSA(green)
1986
1988
1990
1992
1994
1996
1998
-3
-1
0
1
2
3

Figure 12 shows bubble plots for the residuals of numbers-at-age of landings, discards and the two tuning fleets. Tracking cohorts is not very clear. It looks there is no year effect. In landings higher residuals are for ages 0 and 7+ and in Coruña trawl since 1994 almost all values are negative, in opposite with medium years where values are positive.



As in the model for southern hake, when discards are incorporated, recruitment and fishing mortality for young ages increase. We need to evaluate the influence on biological reference points and make projections to detect short and long-term effects on the assessment.



References:

Fernández, C., Cerviño, S., Pérez, N., and Jardim, E. 2010. Stock assessment and projections incorporating discard estimates in some years: an application to the hake stock in ICES Divisions VIIIc and IXa. ICES Journal of Marine Science, 67: 1185–1197.

Working Document 6

Separating catches into landing and discards in the state-space assessment model used for North Sea Cod

Anders Nielsen

Introduction

For a number of years the North Sea Cod assessment has included the somewhat controversial estimation of "unallocated mortality". The unallocated mortality was introduced, because a mismatch was observed between the signal from the survey(s) and the signal from the commercial catches. The root cause for this mismatch could be black landings, wrong estimates of discards, wrongly specified natural mortality, changes in how well the survey(s) covered the stock, or other issues. In the two different assessment models used (Badapt and more recently the State-space Assessment Model SAM) this unallocated mortality has taken the form of separate yearly scaling factors S_y applied to all the total catch- at-age from 1993 and onwards, such that the catches used by the assessment model becomes $S_y C_{a,y}$. In more recent years, information from national authorities indicates that the level of misreporting (referring to landings rather than the whole catch) has been decreasing and is likely to have become negligible since about 2006 (ICES 2011). The landings since 2006 could therefore be considered unbiased, the landings can be considered unbiased, such that an observed mismatch between catches and survey is likely to be caused by wrongly estimated discards.

In this section the currently applied assessment model is summarized, and two approaches for applying the scaling constant solely to the discards are investigated.

Summary of the state-space model for North Sea Cod

The state-space assessment model contains two parts. The first part describes the process of underlying unobserved states α, which are the log-transformed stock sizes $\log N_1, \ldots, \log N_A$ and fishing mortalities $\log F_{i_1}, \ldots, \log F_{i_n}$. The transition equation (below) describes the distribution of the next year's state from a given state in the current year.

$$\alpha_y = T(\alpha_{y-1}) + \eta_y$$

The transition function T is where the stock equation and assumptions about stock–recruitment enters the model. For the stock sizes this becomes:

$$\begin{aligned}
\log N_{1,y} &= \log\Big(R\big(w_{1,y-1}p_{1,y-1}N_{1,y-1} + \ldots + w_{A,y-1}p_{A,y-1}N_{A,y-1}\big)\Big) + \eta_{1,y} \\
\log N_{a,y} &= \log N_{a-1,y-1} - F_{a-1,y-1} - M_{a-1,y-1} + \eta_{a,y} \ , \qquad 2 \le a < A \\
\log N_{A,y} &= \log\big(N_{A-1,y-1}e^{-F_{A-1,y-1}-M_{A-1,y-1}} + N_{A,y-1}e^{-F_{A,y-1}-M_{A,y-1}}\big) + \eta_{A,y}
\end{aligned}$$

Here $M_{a,y}$ is the year- and age-specific natural mortality parameter, $w_{a,y}$ is weight in stock, and $p_{a,y}$ is proportion mature, all of which are assumed known from outside sources. $F_{a,y}$ is the fishing mortality. The function R describes the relationship between stock and recruitment (for North Sea Cod a Beverton–Holt curve is assumed). The parameters of the chosen stock–recruitment function are estimated within the model. All noise terms for the logarithm of the stock sizes are assumed independent normal distributed.

The logarithm of the fishing mortalities are assumed to follow random walks. The random walks are allowed to be correlated to mimic the parallel time-series often observed for fishing mortalities in the different age groups. Define $F_y = \left(F_{1,y}, F_{2,y}, \dots, F_{A,y}\right)'$, then it is assumed that

$$\log F_y = \log F_{y-1} + \xi_y \,, \quad \text{where} \quad \xi_y \sim N(0, \Sigma)$$

where Σ is defined via the standard deviation for the individual processes $\sqrt{\Sigma_{i,i}}$ and the common correlation coefficient ϱ , by $\Sigma_{i,j} = \varrho\sqrt{\Sigma_{i,i}\Sigma_{j,j}}$. The correlation coefficient ϱ is estimated within the model, and this structure allows the model to range from independent random walks (values of ϱ near zero) to a multiplicative F pattern (values of ϱ near one).

The second part of the state-space assessment model describes the distribution of the observations x given the underlying states α. Here x consist of the log-transformed catches and survey indices.

The combined observation equation is:

$$x_y = O\left(\alpha_y\right) + \varepsilon_y$$

The observation function O consists of the familiar catch equations for fleets and surveys, and ε_y of independent measurement noise with separate variance parameters for certain age groups, catches, and survey indices. For the logarithm of the survey catches a separate variance parameter is used for the youngest age group and a common one for all older age groups. For the logarithm of the total catches a separate variance parameter is used for each of the two youngest age groups, and a common one for all older age groups. An expanded view of the observation equation, in the case where no accounting for unallocated mortality is done, becomes:

$$\log(C_{a,y}) = \log\left(\frac{F_{a,y}}{Z_{a,y}}(1 - e^{-Z_{a,y}})N_{a,y}\right) + \varepsilon_{a,y}^{(o)}$$
$$\log(I_{a,y}^{(s)}) = \log\left(Q_a^{(s)} e^{-Z_{a,y}\frac{day^{(s)}}{365}} N_{a,y}\right) + \varepsilon_{a,y}^{(s)}$$

Here Z is the total mortality rate $Z_{a,y} = M_{a,y} + F_{a,y}$, $day^{(s)}$ is the number of days into the year where the survey s is conducted, and $Q_a^{(s)}$ are model parameters describing the catchabilities. Finally $\varepsilon^{(o)} \sim N(0, \sigma_{o,a}^2)$ and $\varepsilon^{(s)} \sim N(0, \sigma_{s,a}^2)$ are all assumed independent and normally distributed.

The likelihood function for this is set up by first defining the joint likelihood of both random effects (here collected in the α_y states), and the observations (here collected in the x_y vectors). The joint likelihood is:

$$L(\theta;\ \alpha;\ x) = \prod_{y=2}^{Y} \{\phi(\alpha_y - T(\alpha_{y-1}), \Sigma_\eta)\} \prod_{y=1}^{Y} \{\phi(x_y - O(\alpha_y), \Sigma_\varepsilon)\}$$

Here θ is a vector of model parameters, and ϕ is the density for a Normal distribution. Since the random effects α are not observed inference should be obtain from the marginal likelihood:

$$L_M(\theta;\ x) = \int L(\theta;\ \alpha;\ x) d\alpha$$

This integral is difficult to calculate directly, so the Laplace approximation is used (via AD Model Builder). The approximation has been verified via two alternative methods. An unscented Kalman filter and importance sampling.

Catch scaling as is currently practiced

The current way to account for unallocated mortality in the assessment model is to apply and estimate a yearly scaling coefficient to the observed catches, from the year 1993 and onwards. This changes the catch-equation above to:

$$\log(C_{a,y} S_y) = \log\left(\frac{F_{a,y}}{Z_{a,y}}(1 - e^{-Z_{a,y}}) N_{a,y}\right) + \varepsilon_{a,y}^{(o)}$$

where S_y is fixed to 1 for years prior to 1993, and estimated thereafter.

It is somewhat unusual to have model parameters on the left hand side of the observation equation like this, but in this case it is unproblematic, as it can easily be rearranged as:

$$\log(C_{a,y}) = \log\left(\frac{F_{a,y}}{Z_{a,y}}(1 - e^{-Z_{a,y}}) N_{a,y}\right) - \log(S_y) + \varepsilon_{a,y}^{(o)}$$

This model is what is currently used for the assessment of North Sea Cod.

Discard scaling, first crude approximation

The catch-at-age data $C_{a,y}$ used for the assessment is the sum of two components The landings-at-age $L_{a,y}$ and the estimated discards $D_{a,y}$. It is plausible that the precision, and in this context more interestingly, the bias of these two data sources are different.

According to national authorities, levels of misreported landings have been negligible since around 2006 (ICES 2011), so that the bias in landing could be considered negligible from 2006 onwards, such that an observed mismatch between catches and survey is likely to be caused by a bias in the estimated discards. A first approach is to change the model formulation to:

$$\log(L_{a,y} + D_{a,y}S_y) = \log\left(\frac{F_{a,y}}{Z_{a,y}}(1 - e^{-Z_{a,y}})N_{a,y}\right) + \varepsilon_{a,y}^{(o)}$$

With the goal of multiplying the scaling S_y to the different selectivity in the discards, instead of the selectivity of the total catch.

There are however a few things to notice about this approach. Firstly, notice that it is not easily possible to collect the model parameters on the right hand side of the observation equation, and the data on the left. This may not be very important, as it is still possible to write down the likelihood, optimize it, and obtain estimates of our model parameters. It does imply a correlation between the estimated scaling parameters and the estimated variance of the catch observations, but correlated estimates are not unusual in non-linear models. Secondly, notice that the fishing mortality estimated in this model $F_{a,y}$ will correspond to the total catch, which is comparable to the currently applied model. Finally, notice that if discards estimates are not known with the same precision as landings-at-age, then the model should have separate variance parameters for the two catch components, and this model cannot. This model is however presented as a simple alternative to the following.

Discard scaling, by splitting landings and discards

Instead of modelling the sum of landings and discards $C_{a,y} = L_{a,y} + D_{a,y}$ it may be more reasonable to model the two data sources directly. Such a model requires a more substantial change in the model. The observation equation is extended from one for catches into one for landings and one for discards:

$$\log(L_{a,y}) = \log\left(\frac{F_{a,y}^{(L)}}{Z_{a,y}}(1 - e^{-Z_{a,y}})N_{a,y}\right) + \varepsilon_{a,y}^{(L)}$$
$$\log(D_{a,y}S_y) = \log\left(\frac{F_{a,y}^{(D)}}{Z_{a,y}}(1 - e^{-Z_{a,y}})N_{a,y}\right) + \varepsilon_{a,y}^{(D)}$$

with total mortality now redefined as $Z_{a,y} = M_{a,y} + F_{a,y}^{(L)} + F_{a,y}^{(D)}$. Notice that this model has separate fishing mortalities and different noise terms (with separate variance parameters) for landings and for discards.

The unobserved random processes of the state-space model also need to be altered. Instead of having processes for $\log N_{a,y}$ and $\log F_{a,y}$, we now need to have processes $\log N_{a,y}$ and for two sets of fishing mortalities $\log F_{a,y}^{(L)}$ and $\log F_{a,y}^{(D)}$. Independent random walks are assumed for the logarithm of the fishing mortalities corresponding to the discards. This allows the selectivity of the landings and the selectivity of the discards to develop independently.

Theoretically this model is well suited to handle the task of applying the unallocated mortality to the discards only, but there is one important practical concern. Discarding is mainly done in age classes one and two; at older ages the number of discarded fish are frequently zero. This causes a problem for the split-model described here, as the logarithm transformation is not possible ($\log(0) = -\infty$). Dealing with a few zeroes is fairly common in assessment models, and 'solutions' often seen are:

1) adding a small constant to all zeros,

2) ignoring those observations, or

3) setting up a different likelihood that allows for zeros, which could for instance be a negative binomial.

Here the problem is not a few zeros (Figure 1), but an entire age group or long periods of zeros. Having long periods of constant zero catch also corresponds poorly with the random walk assumption for $\log F^{(D)}$ in the model. To get something operational within the time limits of WGMG only age groups one and two were included for discards in this model. In older age groups, only a very small fraction of the total catch is discarded, except for in most recent years (Figure 1).

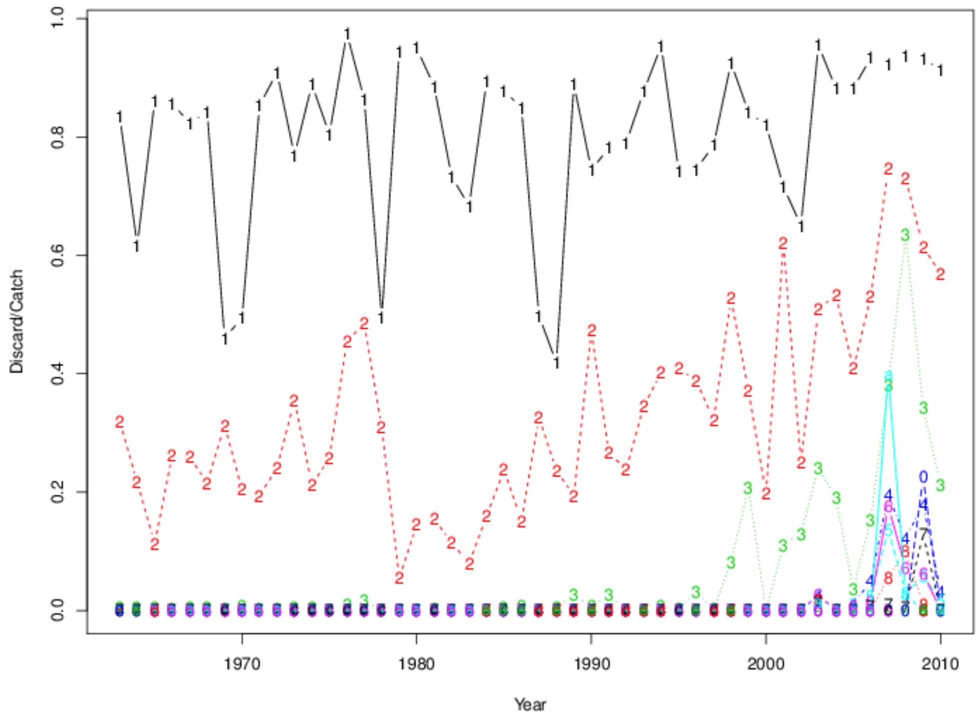


Figure 1. The fraction of the total catch discarded in each age group.

Results

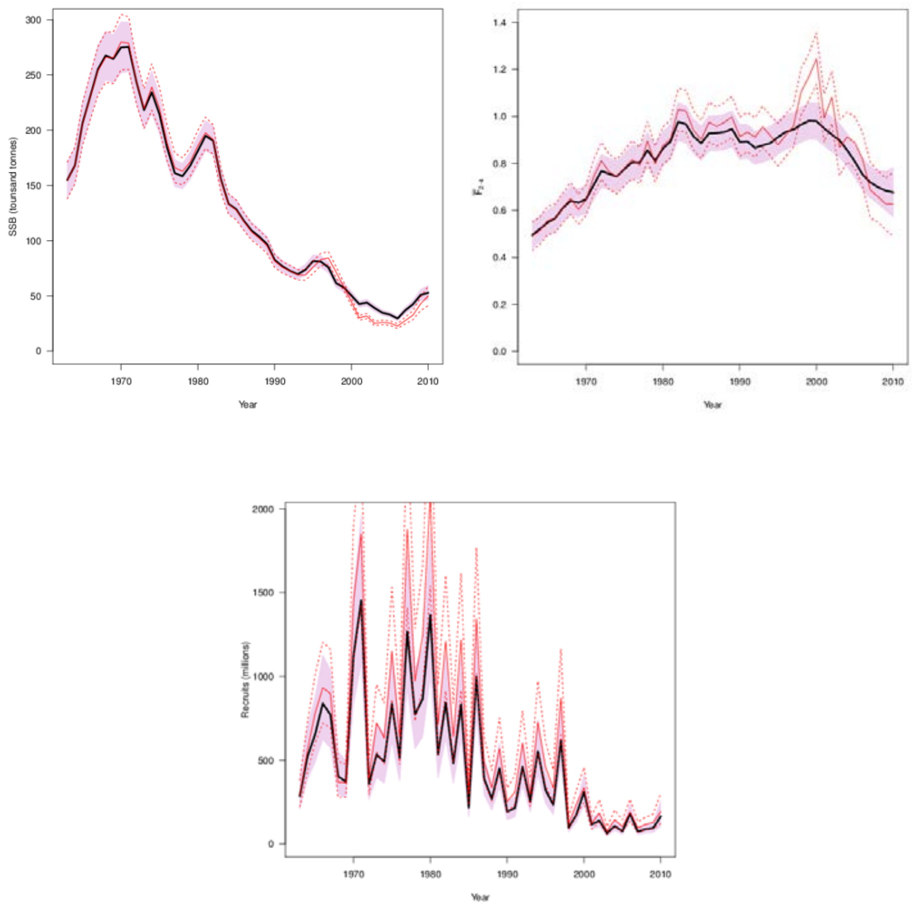


Figure 2. Comparing the catch scaling currently applied (black solid line with shaded confidence interval) to results from the split model (solid red line with red dashed confidence intervals) where scaling is only applied to the discarded part.

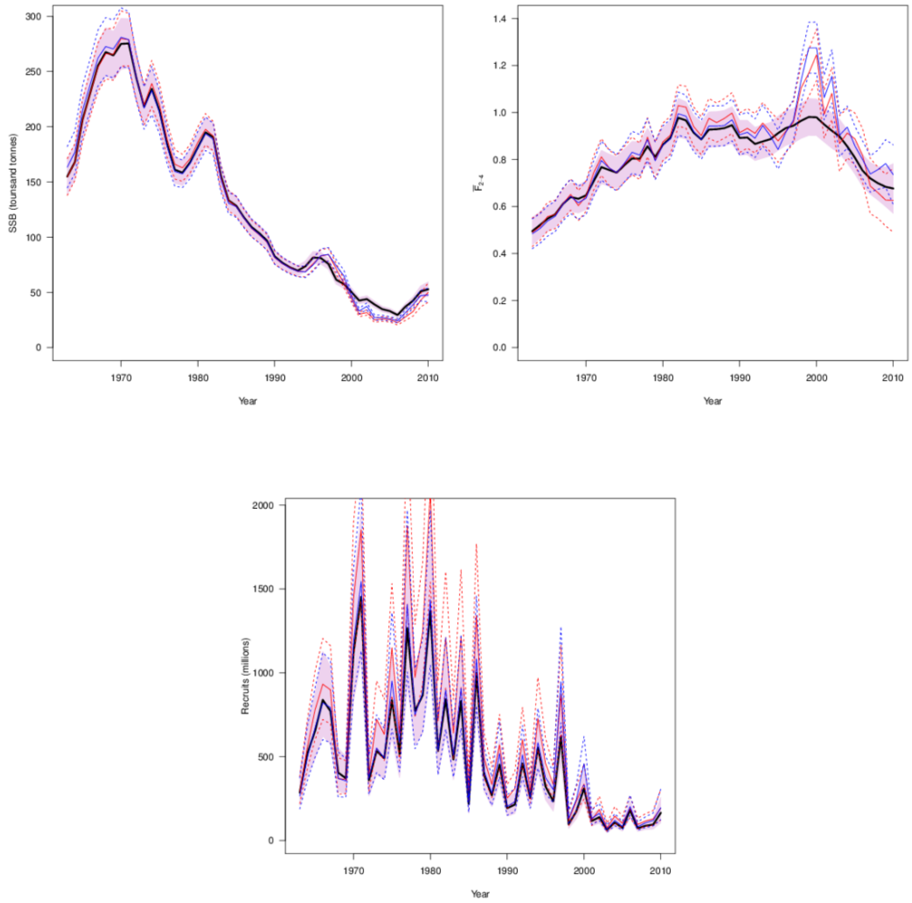


Figure 3. Catch scaling currently applied (black solid line with shaded confidence interval), results from the split model (solid red line with red dashed confidence intervals) where scaling is only applied to the discarded part, and results from the simple approximation where data are not split, but scaling is applied to the selectivity implied by discards (solid blue line with blue dashed confidence intervals).

Table 1. The likelihood values for comparing the different configurations for the two different approaches.

Model	−logL (split)	−logL (simple)	Number of parameters
Scaling on catch	199.260	95.99	36/34
Scaling on landings	199.212	94.52	36/34
Scaling on discards	233.336	115.03	36/34

Conclusions

It would be useful if the assessment working group could pin down more precisely where the mismatch between survey and catches comes from. It has been demonstrated that the state-space assessment framework is flexible enough to allow such additions to the model.

Two approaches have been developed: a simple approximation, which retains catch (the sum of landings and discards) as the observed variable in the model, but applies the unallocated mortality to the discarded part only, and a more complex model, which separates the observed variables into landings and discards. Both approaches showed the same stock trends, and gave the same conclusions w.r.t. the question about origin of the unallocated mortality.

The simple approach seemed to work as well in this study, as the more complex splitting approach. The only difference seen in the results is a modestly higher recruitment for the splitting model in the first part of the data period. It is expected that the first age class is the most sensitive to splitting the catch into landings and discards, since discarding is mainly done for the youngest ages. The higher recruitment in the first part of the data period could be due to the splitting, thereby estimating different selectivity patterns, or (perhaps more likely) due to the fact that for the splitting model, it was necessary to remove discards for ages 3 and older. This restriction was not needed in the simple approach. In terms of running time the simple approach is also preferable.

The model for which unallocated mortality over the entire period (from 1993 to 2011) is assigned to the discard part only fitted significantly worse than the model where the unallocated mortality is assigned to the catch. Notice that this conclusion is for the entire period with discard scaling compared to the entire period with catch scaling.

WGMG focused on the development of the methods to handle these comparisons, not on formulating and testing the more specific interesting hypothesis, relating to the timing of different regulatory measures. It is recommended that the assessment working group use these developed methods to test such hypothesis.

References

ICES. 2011. Report of the Workshop on the Analysis of the Benchmark of Cod in Subarea IV (North Sea), Division VIId (Eastern Channel) and Division IIIa (Skagerrak; WKCOD 2011), 7–9 February 2011, Copenhagen, Denmark. ICES CM 2011/ACOM:51. 94 pp.

Working Document 7

An initial comparison of the performances of simple management procedures compared to complex assessments for some ICES stocks

H.F. Geromont and D.S. Butterworth

Marine Resource Assessment and Management Group,
Department of Mathematics and Applied Mathematics
University of Cape Town, Rondebosch 7701, South Africa

Abstract

These analyses aim to compare the fishery and resource consequences of management recommendations based on complex annual resource assessments to those based on simple empirical management procedures (MPs), which in the cases considered use only the annual abundance estimates from a single survey. The 2010 ICES assessments of the stocks of North Sea Plaice and Sole in Subarea IV are used for the investigation. The MPs are selected from the results of simulations based only on the resource information available in 1990. Their performances are then compared to what actually transpired over the 1990 to 2009 period under advice arising from the regular ICES assessments. For plaice, almost without exception the MPs' performances dominate what actually eventuated for every performance statistic: higher catches, greater final spawning biomass, lesser lowest spawning biomass during the 20 years, lower average fishing mortalities, and lesser inter-annual variation in both catch and fishing mortality. For Sole these results are qualitatively duplicated, except for marginally smaller catches in some cases. In circumstances for ICES stocks where there may be difficulties in sustaining the level of sampling required for complex annual assessments, such as annual ageing of the catch, because of diminishing resources, these results are sufficiently promising to suggest that they be extended, in particular to further stocks, to confirm whether they might indeed provide an defensible alternative approach to the provision of scientific management advice.

1 Introduction

ICES Working Groups have generally based scientific management advice for more valuable stocks on regular (often yearly) assessments. These stock assessments using age-structured population models such as ADAPT-VPA, XSA and SAM, are often very complex and require a substantial amount of expertise, time and effort. The assessment process is further complicated by decisions regarding which data to include in the analyses, and how these data are incorporated in the objective function being minimized to fit the population model.

However a detailed annual stock assessment may not be necessary to achieve management goals and may constitute an inappropriate use of limited resources. Is there a simpler, more efficient, way of providing reliable management advice? This question is all the more relevant at this time, with diminishing resources raising questions in ICES over whether annual ageing of catches required for assessment methods such as XSA can be sustained.

This paper performs initial analyses to investigate whether simple empirical management procedures (MPs) might perform as well as these complex annual assessments in achieving management goals. This approach, also known as Management Strategy Evaluation (MSE), has established itself as a powerful fisheries management tool to assist meet multiple management objectives in a manner that checks robustness to uncertainty for compatibility with the Precautionary Approach (De Oliveira *et al*., 2010). The 2010 ICES assessments of the stocks of North Sea Plaice and Sole in Subarea IV (ICES WGNSSK Report 2010) are used as the basis to compare these two management approaches. The comparison consists of four steps.

I) *Deterministic "hindsight" projections*

Three simple empirical MPs are each tuned to achieve over the last 20 years (1990 to 2009) the same final (2009) spawning biomass as estimated by the assessments. Since these projections may each follow a different trajectory to that suggested by the assessments, some assumptions are needed to be able to effect the computations.

i) In allocating the annual catch among the different ages, the same selectivity-at-age as estimated in the assessment for the year concerned is assumed to apply.

ii) If spawning biomass differs from that in the assessment, recruitment would presumably do so too. A stock–recruitment relationship is fitted to the output from the assessment, with an associated multiplicative residual calculated for each year to reflect the difference between the actual (assessment) recruitment for that year and the value suggested by the stock–recruitment relationship fitted. In these "hindsight" projections, for which the spawning biomass in a particular year may differ from that in the assessment, the key assumption made is that the ***same*** multiplicative residual will apply to the expected recruitment calculated from the estimated stock recruitment relationship. Thus, for example, if the recruitment indicated by the assessment for 1996 was 20% above the value suggested by the fitted stock–recruitment relationship, in any other projection this same 20% will be added to the recruitment predicted by the stock–recruitment relationship for that year, given the projected spawning biomass that year.

iii) Some of the MPs make use of the annual abundance estimates from a survey. Those estimates will differ from the results expected from the best fit of the assessment model to these data in the assessment process. As in ii) above, the assumption made for the projections, for which the underlying abundance under an MP may differ from that for the assessment itself, is that the multiplicative residual for the assessment applies also to the survey estimate which would have resulted under the MP for the year concerned.

These MPs with their associated tunings are referred to as "hindsight" MPs, as they have the benefit of hindsight in "knowing" what will happen in the next 20 years in terms of uncertainties (residuals related to recruitments and survey sampling errors.

II) *Stochastic "forecast" projections of "hindsight" MPs.*

If one had been choosing an MP twenty years ago, one would not have had the benefit of the "hindsight" above at that time. Rather than knowing exactly what recruitment residual will apply each year in future, projections have to assume that these will be drawn at random each year from distributions estimated from fits of stock–recruitment relationships to the assessment results available at that time (which are taken here to be the 2010 assessment results up to 1989). Similar assumptions need to be made about the abundance estimates forthcoming from future surveys.

When the "hindsight" MPs are applied under these stochastic "forecast" conditions, rather than with exact knowledge of the future, their performance deteriorates, in particular in often yielding final biomasses after 20 years which are considerably below those which actually eventuated. The purpose of this step is to check whether the performance of these "hindsight" MPs would have been considered sufficiently acceptable to have led to their implementation 20 years ago.

III) *Use of stochastic "forecast" projections to tune MPs*

This step involves selecting alternative tunings of the three simple empirical MPs considered in step I) that might have led to their being considered acceptable 20 years ago. The stochastic projections are used to select control parameters for these MPs that achieve a spawning biomass distribution in 20 years time which at the lower 2.5% level is at least as large that which the assessment indicates to have actually resulted. These more conservative MPs are termed "forecast" MPs.

IV) *Performance of selected "forecast" MPs under "hindsight" projections*

In this final step, the "forecast" MPs selected at step III) are applied in conjunction with the deterministic "hindsight" projections (the residuals that actually "occurred") to determine how well those MPs would have managed the fisheries considered. The fundamental question to be addressed is how do the resultant averages catches and fishing mortalities, their interannual variability, and the final spawning biomass after 20 years compare to what was achieved in practice under management based on the use of advice arising from annually updated assessments.

2 Technical specifications of projections

The projection period spans the last twenty years of the 2010 ICES assessments, i.e. from 1990 to 2009. Therefore, projections commence in 1990 and are moved forward year by year by first obtaining the TAC according to a particular MP based on new survey biomass data, from which the corresponding fishing mortality, F_y, rate can be computed for that year given the selectivity-at-age vector selected. The population numbers for the next year can then be computed. The number of recruits (1-year olds) for the next year is then calculated using a Beverton–Holt stock–recruit relationship (see Appendix A for specifications of assessment data and parameters used in the projections).

Population numbers-at-age are projected forward from 1990 to 2009 using the following equations which assume continuous fishing throughout year (Baranov equation):

$$N_{y+1,a\min} = R_{y+1} \tag{1}$$

$$N_{y+1,a+1} = N_{y,a} e^{-(M_a + S_{y,a} F_y)} = N_{y,a} e^{-Z_{y,a}} \quad \text{for } a_{\min} \le a < m-2 \tag{2}$$

$$N_{y+1,m} = N_{y,m-1} e^{-(M_{m-1} + S_{y,m-1} F_y)} + N_{y,m} e^{-(M_m + S_{y,m} F_y)} \tag{3}$$

where

$N_{y,a}$ is the number of fish of age *a* at the start of year *y*,

M_a denotes the natural mortality rate on fish of age *a*, which is input,

$S_{y,a}$ is the age-specific selectivity for year *y*, which is input (deterministic) or randomly sampled (stochastic),

F_y is the fishing mortality for year *y*, which is estimated,

$m = 10$ is the maximum age considered (taken to be a plus-group), and

$a_{\min} = 1$ is the minimum age considered.

Stock–recruitment relationship:

The "future" number of recruits at the start of year *y* from 1990 to 2009 is related to the spawning stock size by a stock–recruitment relationship. Two forms of such a relationship are considered. The first is the Beverton–Holt form

$$R_y = \frac{\alpha \, B^{sp}_{y-1}}{\beta + B^{sp}_{y-1}} e^{\varsigma_y} \tag{4}$$

where

α and β are the stock–recruitment parameters estimated by minimizing the negative of the log likelihood in equation (A.13) of Appendix A, which are input,

ς_y are the corresponding recruitment residuals which are either input for the deterministic projections, or $\varsigma_y \sim N\left(0,(\sigma^R)^2\right)$ for the stochastic projections with standard deviation of $\sigma^R = 0.5$ for both stocks (this value was used because it is close to the estimated standard deviation and for the sake of simplicity), and

B^{sp}_{y-1} is the spawning biomass in year $y-1$, given that

$$B^{sp}_y = \sum_{a=1}^{10} f_a w^S_{y,a} N_{y,a} \tag{5}$$

where

$N_{y,a}$ is the projected number of fish in year y of age a given by equations (1), (2) and (3), and

$w^S_{y,a}$ are the population weights-at-age for each year used in the 2010 ICES assessments, and f_a is the proportion of fish of age a that are mature, which is input.

The second is a two-line (or "hockey stick") form

$$\begin{aligned} B^{sp}_y \geq B^0: \quad & R_y = \alpha e^{\zeta_y} \\ B^{sp}_y < B^0: \quad & R_y = (\alpha B^{sp}_{y-1} / B^0) e^{\zeta_y} \end{aligned} \tag{6}$$

where B^0 is the minimum spawning biomass over the period under consideration, and α is estimated as above (which will yield the geometric mean of the recruitment estimates over this period).

Catch equation:

Once a TAC for year y is generated by the MP, the corresponding fishing mortality rate, F_y, can be computed. When using the Baranov formulation, the total number of fish caught of age a in year y is given by

$$C_{y,a} = N_{y,a} \frac{S_{y,a} F_y}{Z_{y,a}} (1 - e^{-Z_{y,a}}) \tag{7}$$

Where

F_y is computed using the bisection method such that

$$C_y = \sum_{a=1}^{10} w_{y,a}^C \, C_{y,a} \tag{8}$$

where

C_y is the total annual catch (TAC) corresponding to a chosen harvesting strategy, and

$w_{y,a}^C$ are the catch weights-at-age for each year taken from the 2010 ICES assessment, which are input.

Survey biomass:

The future biomass corresponding to survey index *i* is given by

$$B_y^{sur-i} = \sum_{a=1}^{10} S_a^{sur-i} w_{y,a}^S N_{y,a} \tag{9}$$

where

$w_{y,a}^S$ denote the population weights-at-age for each year used for the 2010 ICES stock assessment, which are input, and

S_a^{sur-i} is the fishing selectivity corresponding to the survey index *i*.

Projected survey data

Future survey data are generated assuming the same residuals as inferred from the adjusted 2010 XSA assessment

$$I_y^i = q^i B_y^{sur-i} e^{\varepsilon_y^i} \tag{10}$$

where

B_y^{sur-i} is the model estimate of projected survey biomass, given by equation (9),

q^i is the constant of proportionality for survey abundance series *i* estimated using equation (A.19) in Appendix A, and

ε_y^i are the residuals given by equation (A.20) in Appendix A for the deterministic projections, or $\varepsilon_y^i \sim N\left(0, (\sigma^i)^2\right)$ for the stochastic projections, where σ^i are either given by equation (A.21), or input. For these projections the standard deviation was fixed to $\sigma^i = 0.2$ for both stocks for simplicity, the value being approximately equal to the estimated standard deviation.

Projected commercial selectivity:

The commercial selectivity-at-age vectors for future years (1990 onwards) are sampled randomly from past (1980 to 1989) XSA estimates.

Projected weights:

The population, $w^S_{y,a}$, and catch, $w^C_{y,a}$, weights-at-age for future years (1990 onwards) are set equal to the average weight for each age over the last three years prior to the projection period, i.e.

$$w_{y,a} = 1/3 \sum_{y'=1987}^{1989} w_{y',a} \text{ for y>1989.}$$

3 Candidate Management Procedures

Some very simple empirical management procedures, based on trends in survey indices of abundance, are investigated. These simple MPs are particularly useful in data-poor situations where data are limited (Geromont and Butterworth, 2010), rendering a model-based MP unsuitable, or where there is too much variability about the data, in which case a more complex model-based MP may well follow noise rather than trend. Furthermore, the very simple empirical rules are easy to understand, test and apply and have been shown to be as robust to uncertainty as their model-based counterparts in a number of cases (for example in the development of MPs for Southern Bluefin Tuna – CCSBT, 2010). The main disadvantage of empirical MPS are that there are no estimates of resource abundance and other management reference points on which to base TACs.

For example "derivative" or "slope"-based MPs utilize the trend in a limited subset of data (typically the most recent 5 years of survey biomass estimates) for input. The annual TAC is simply moved up or down from where it was the previous year without knowledge of where the resource might be in relation to maximum sustainable yield level or other conventional management reference points. Note that in implementation for relatively data-rich stocks such as North Sea Plaice and Sole that are considered here, a simple MP approach like this would be underpinned by a full resource assessment; the former provides ongoing yearly management advice, while the latter is re-considered at multiyear intervals to re-check the appropriateness of the MP and if necessary to adjust some of its parameters.

3.1 Constant catch MP

At the one extreme, this is the simplest of all empirical MPs and requires no data to set the annual TAC. The future TAC given by

$$TAC_{y+1} = TAC^{target} \tag{11}$$

where TAC^{target} is chosen such that the projected population spawning biomass in 2009 reaches some target level $B^{sp}_{2009} = B^{target}$. For the "hindsight" projections, the target biomass was chosen to be equal to the final spawning biomass estimated in the adjusted 2010 ICES assessment, $B^{target} = B^{XSA}_{2009}$, to facilitate comparison between the MP and assessment-based management approaches.

For the stochastic "forecast" projections, a search routine is used to find the constant catch that reaches the target for each simulation, an approach suggested by Bentley and Langley (2011). The desired constant level for future catches is selected from the resulting distribution as the one that will provide adequately risk adverse performance under the uncertainty incorporated in the projections (the 2.5%-ile value was chosen for these projections).

Note: A constant catch harvesting strategy is not recommended as there is no feedback-control mechanism built into this type of MP. It does however give a ball-park figure of the average yield that can be expected during the projection period given a chosen target, which is useful for later comparison of the different candidate MPs.

3.2 Survey slope based MP

For this type of MP, limited data are used in the MP formula to ascertain recent trends in biomass, with the TAC being moved up or down depending on whether the perceived trend is positive or negative. The TAC for the next year is given by

$$TAC_{y+1} = TAC_y(1+\lambda s_y) \tag{12}$$

Where

TAC_y is the TAC in year *y*,

λ is a control parameter that reflects how strongly the TAC is adjusted in response to the perceived trend in resource biomass, and

s_y is a measure of the trend in the survey abundance index given by the slope of the linear regression of $\ln I^i_{y'}$ against y' for years $y' = y-p, y-p+1, ..., y$ for abundance index I^i, and

p is the number of years over which the slope is calculated. Note that if p is too small the trend estimates would fluctuate too much (tracking noise) and if p is too large the MP would not be able to react quickly enough to recent trends in resource abundance. A value

For the first year of the projection period an appropriate "starting level", TAC^*, must be chosen (not necessarily equal to the actual TAC that year) . This is specified as $TAC^* = x\%TAC_n$, where *n* is the last year of the assessment period and *x* is a control parameter that reflects how aggressive/conservative the MP should be. The choice of this starting point is important for the performance of the MP because a starting level that is too low will result in an unrealistically large drop in TAC in the first year of management (unrealistic because it would not be accepted in practice), while a starting point that is too high necessitates subsequent severe cuts in the TAC.

In addition, with the exception of the starting level TAC, TAC_{n+1}, all subsequent TACs are restricted to increase/decrease by at most *v*% from the previous year, i.e. let

$$TAC_{change} = (TAC_{y+1} - TAC_y) / TAC_y \tag{13}$$

then

$$TAC_{y+1} = TAC_y + v\%TAC_y \qquad \text{if } TAC_{change} > v\%$$

or

$$TAC_{y+1} = TAC_y - v\%TAC_y \qquad \text{if } TAC_{change} < -v\%$$

A restriction of 20% interannual variation in catch was chosen for base case runs in order to be reasonably consistent with the maximum annual changes in observed catches (landings plus discards).

3.3 Target based MP

This type of MP is based on moving resource abundance to a chosen target level for some abundance index *I*. The form of the Tier 4 control rule in Wayte (2009) is used here. The TAC is adjusted up or down depending whether the most recent abundance index (in these cases survey biomass estimate) is above or below the target survey.

The TAC for the next year is given by

$$TAC_{y+1} = TAC^{target}[w + (1-w)(\frac{I^{recent} - I^0}{I^{target} - I^0})] \tag{14}$$

if $I^{recent} \geq I^0$ and

$$TAC_{y+1} = wTAC^{target}\left(\frac{I^{recent}}{I^0}\right)^2 \tag{15}$$

if $I^{recent} < I^0$

where

I^{recent} is the average survey over the most recent four years,

$I^{target} = x\%I^{ave}$ is the desired target value for the survey index of abundance,

$I^0 = y\%I^{ave}$ is a lower survey abundance index level below which the TAC decreases to zero rapidly,

$I^{ave} = 1/5\sum_{y=1985}^{1989} I_y$ is an average historic survey abundance index value,

TAC^{target} is the catch target (when $I^{recent} = I^{target}$), and

w is a fraction that defines the catch level when $I^{recent} = I^0$.

A simplified, commonly used, form of equation (14) is obtained by setting *w=0*

$$TAC_{y+1} = TAC^{target}\left(\frac{I^{recent} - I^0}{I^{target} - I^0}\right) \quad (16)$$

Here, the catch is set to zero when the abundance index reaches its lower limit, I^0. At the other extreme, setting $w=1$ results in the constant catch harvesting strategy of equation (11).

However, the formulation given by equation (14) allows for a non-zero catch of $wTAC^{target}$ when $I^{recent} = I^0$, which has the effect of dampening the interannual variation in catches, thereby stabilizing the output from the MP. Setting $w=0$ would necessitate a steeper slope of the linear relationship given by equation (16), leading to more variable catches. On the other hand, setting $w=1$ would result in no interannual fluctuations in catch, but also no adjustment of catch in response to changes in survey abundance indices. A suitable trade-off between the level of feedback control and interannual catch variation was sought and a value of $w=0.5$ was chosen for the deterministic retrospective projections considered here, so that equation (14) becomes

$$TAC_{y+1} = 0.5TAC^{target}\left[1 + \left(\frac{I^{recent} - I^0}{I^{target} - I^0}\right)\right] \quad (17)$$

In addition, a restriction for maximum allowed interannual change in catch is imposed as per equation (13).

Figure 1 illustrates these forms of relationship for different values of the control parameter *w*.

4. Results

The statistics reported for comparison of performance of the MPs over the period from 1990 to 2009 are:

i) average annual catch over the projection period, $\overline{TAC}$,
ii) average annual variation of catch (variation given by modulus of change in catch as a proportion of previous catch) over the projection period, $\overline{\Delta TAC}$,
iii) the average annual fishing mortality rate, $\overline{F}$,
iv) the average annual variation (given by modulus of change) in fishing mortality, $\overline{\Delta F}$,
v) minimum spawning biomass as a fraction of the target biomass, $\min B_y^{sp} / B^{target}$ where B^{target} corresponds to the 2009 spawning biomass estimated in the adjusted 2010 ICES assessments for North Sea Plaice or Sole in Subarea IV, and
vi) final (2009) spawning biomass expressed in the same way as detailed in v).

Three types of MPs with different harvest control rules are investigated:

- Constant catch MP, i.e. no data used in the TAC setting rule (equation (11)).

- Survey slope based MP, using trend information in the survey index of abundance to increase/decrease TACs from one year to another in accordance to a positive/negative trend (equation (12)).
- Target based MP, in which a target survey estimate is specified and the future TAC is adjusted each year to approach, and eventually maintain, the target (equation (14)).

Two stock–recruit functions are investigated:

- A Beverton–Holt stock–recruit curve with a steepness fixed at 0.9.
- A simple 2-line ("hockey-stick") stock–recruitment relationship.

For North Sea Plaice and Sole in Subarea IV, the survey data used in the TAC setting rules are the BTS-Isis index, aggregated over all ages.

Results are shown in Tables 1 to 4 and Figures 2 to 6 for North Sea Plaice, and Tables 5 to 8 and Figures 7 to 11 for Sole in Subarea IV. Deterministic ("hindsight") and stochastic ("forecast") projections are performed starting in year 1990 to 2009. One thousand simulations were run for the stochastic analyses. Four sets of results are shown for each stock:

- Deterministic "hindsight" projections with MPs tuned to hit the 2009 target spawning biomass exactly (Tables 1a, 1b, 5a and 5b and Figures 2 and 7).
- Stochastic "forecast" projections of "hindsight" MPs (Tables 2a, 2b, 6a and 6b).
- Stochastic "forecast" projections used to tune the MPs: those MPs that showed the best performance while at the lower 5%-ile achieving a final spawning biomass which was equal to or greater than that actually achieved as indicated by the 2010 ICES assessment, were selected and are termed "forecast" MPs (Tables 3a, 3b, 7a and 7b and Figures 3, 4, 8 and 9).
- Deterministic "hindsight" projections, but here under the "forecast" MPs (Tables 4a, 4b, 8a and 8b and Figures 5 and 10).

Throughout, the Tables contrast values of the performance statistics that were actually achieved to those for the three types of MP. Figure 6 provides a graphical summary of the performance statistics for the various tuned "forecast" MPs for North Sea Plaice from projections over 1990 to 2009: (a) shows results for the stochastic "forecast" projections, while (b) shows these for the "hindsight" projection for which the stock–recruitment and survey residuals found in the 2010 ICES assessments are taken to apply. Figure 11 shows the corresponding results for Sole in Subarea IV.

5. Discussion and conclusions

The "hindsight" MPs for which results are reported in Tables 1 and 5 outperform what was achieved in practice for total catch over the 20 year period considered, despite finishing the period with the same spawning biomasses as do the ICES assessments. The constant catch MP performs best in this respect, though the other two types of MP are not far behind. The other performance statistics are nearly always better for the MPs than was achieved historically.

The larger catches with the same target biomass arise in part because the Beverton–Holt stock-relationship with $h = 0.9$ that is assumed reflects an increase in recruitment over the range of spawning biomasses that occur over the 1990–2009 period, so that a harvest pattern that leads to that biomass being nearer the upper end of that range will result in improved recruitment and hence greater productivity than historically.

The actual best estimates of h from the assessment data available up to 1990 are actually higher than 0.9, and the decision to use 0.9 was taken given conventional reluctance among scientists to buy into a relationship that drops below its pristine level only at very low biomass. However, particularly for the North Sea Plaice, this factor contributes to the "larger catch" behaviour mentioned above, and the reason for introducing the 2-line ("hockey-stick") relationship as well was as a form of robustness check in circumstances that would not evidence this feature of greater productivity at larger spawning biomass (as is more consistent with the data).

Despite this good performance, these "hindsight" MPs would not have been viable candidates for implementation in 1990. The reasons are readily evident from Tables 2 and 6. For example, under the stochastic "forecast" projections that take account of future uncertainty in recruitment levels, the median final spawning biomass for North Sea Plaice in 2009 for the "hindsight" constant catch MP is zero. Though for the other "hindsight" MPs as well as for Sole the behaviour is a little better, nevertheless nearly all the MPs exhibit lower 2.5%-iles for spawning biomass that go very low and even to zero.

However once these MPs are tuned under the stochastic "forecast" projections to yield final spawning biomass distributions whose lower 5%-iles are at least as large as occurred in reality, this problem is obviated and the associated "forecast" MPs are certainly such as might have been accepted for implementation in 1990. In terms of medians of performance statistics, these MPs still achieve better performance in nearly all respects (i.e. "dominate") what was achieved in practice for North Sea Plaice (Table 3). For Sole (Table 7) virtually the same applies, the exception being that the average catch is some 10% less, though this comes with the advantage of a final spawning biomass improvement by a factor of about double or more. Tables 4 and 8, where these "forecast" MPs are applied under the "hindsight" projections, show the same patterns.

Figures 6 and 11 provide useful summaries of these performance statistics in graphical form. The domination of the "forecast" MPs performance over actual events for North Sea Plaice is readily evident, though for Sole again the MPs' catch performance is slightly weaker. Of the different types of MPs, the constant catch types reflect the smallest total catches, and in any case are unlikely candidates in reality because of their lack of feedback features to provide robustness to other uncertainties which have not been considered here. The catch performance for the slope and target based MPs is almost equal to what was achieved historically for Sole, and overall there appears little to choose between these two MP types in terms of performance.

In general these results are similar to those obtained in simulation studies by Punt (1993), which showed that compared to simpler management approaches based on production models, attempts to take age-structure information into account through VPA in recommending catch limits led to greater variability in those limits without any corresponding enhancement of performance in terms of resource conservation.

In conclusion, these results seem sufficiently promising to suggest that they be extended, in particular to further stocks, to confirm whether they might indeed provide a defensible alternative approach to the provision of scientific management advice. At present for ICES stocks, because of diminishing research resources, there may be difficulties in sustaining the level of data input required for complex annual assessments, such as annual ageing of the catch. This raises the question of whether such complex assessments can continue to serve as the primary basis to provide scientific advice on catch limits, so that there is a need to explore alternative possibilities as done here.

This is not to suggest that complex assessments can be abandoned. Rather they still need to be conducted from time to time to provide the updated representations of the underlying resource dynamics that serve as the basis for repetitions of the process of re-selecting simple MPs at regular intervals. Though VPA assessments require annual ageing, if that cannot be continued, various SCAA/integrated analysis assessment approaches which do not require age data for every year could be used instead.

6. Future Work

- Due to limited time, only two stocks (North Sea Plaice and Sole in Subarea IV) have been investigated thus far (and the parameter choices for the Sole MPs could not be optimized to the same extent as those for Plaice). The highest priority would seem to be the extension of this work to consider more ICES stocks, hopefully to confirm more widespread applicability.
- In comparing performance above, "forecast" MPs were tuned to achieve the same final spawning biomass level at some low percentile (5%) of the distribution of this statistic. Trade-off comparisons might be more readily made if instead tuning was effected to achieve the same total catch over the period.
- The stochastic projection trial exercise should be extended to incorporate more checks of robustness. Aspects to be considered for inclusion in such an extension include first estimation and then model structure uncertainty in the numbers-at-age vector that commences the projections, variability in natural mortality, and a greater number of stock–recruitment relationships.
- In the calculations above, the TAC specified by the MP was assumed to be exactly equal to total removals for the year concerned. Realistic levels of implementation error need to be incorporated into projections.
- At a later stage, if this approach finds wide favour, rather than demonstrations of adequacy based on history, the analyses will need to move on to consider simulations projecting forward from the present time, so as to develop MPs that can be seriously considered for implementation. This could involve extension beyond the simple types of MPs considered here, as well as a wider range of robustness testing.

Acknowledgements

We thank José De Oliveira, Charlie Edwards and Laurie Kell for assistance in providing the assessment data we have used. Financial support of the National Research Foundation (NRF) of South Africa is gratefully acknowledged.

References

Bentley, N., and A. Langley. 2011. Feasible stock trajectories: a flexible and efficient sequential estimator for use in fisheries management procedures. Canadian Journal of Fisheries and Aquatic Sciences. 2011–0259.R1 In press.

Commission for the Conservation of Southern Bluefin Tuna. 2010. Report of the Scientific Committee: 119p.

De Oliveira, J. A. A., Kell, L. T., Punt, A. E., Roel, B. A., and Butterworth, D. S. 2008. *In* Managing without best predictions: the Management Strategy Evaluation framework. In Advances in Fisheries Science. 50 years on from Beverton and Holt. Payne, A.I.L., Cotter, J.R. and Potter, E.C.E., pp. 104–134. Blackwell Publishing, Oxford. xxi + 547 pp.

Geromont, H. F., and D. S. Butterworth 2010. Initial investigation of generic management procedures for data-poor fisheries. MARAM IWS/DEC10/DPA/P2: 1–41. www.mth.uct.ac.za/maram/pub/2010/MARAM_IWS_DEC10_DPA_P2.pdfICES WGNSSK Report 2010. Report of the Working Group on the Assessment of Demersal Stocks in the North Sea and Skagerrak (WGNSSK). International Council for the Exploration of the. ICES CM 2010/ACOM:13: p 422–491.

Punt, A. E. 1993. The comparative performance of production-model and ad hoc tuned VPA based feedback-control management procedures for the stock of Cape hake off the west coast of South Africa. In Risk Evaluation and Biological Reference Points for Fisheries Management, pp. 283–299. Ed. by S. J. Smith, J. J. Hunt, and D. Rivard. Canadian Special Publication of Fisheries and Aquatic Sciences, 120.

Wayte, S. E. (ed.) 2009 Evaluation of new harvest strategies for SESSF species. CSIRO Marine and Atmospheric Research, Hobart and Australian Fisheries Management Authority, Canberra. 137 p.

PLAICE: SUBAREA IV		BEVERTON–HOLT STOCK RECRUIT RELATIONSHIP WITH H=0.9		
		MP1	MP2	MP3
	Observed catches	Constant catch	MP slope: BTS-Isis $x=80\%, \lambda=0.556$ $p=4$ (max 20% ΔTAC)	MP target: BTS-Isis $I^0=0.2I^{ave}, I^{target}=1.1I^{ave}$ $TAC^{target}=206700$ $w=0.5$ (max 20% ΔTAC)
$\overline{TAC}$	145702	185989	152206	156460
$\overline{\Delta TAC}$	0.166	0.016	0.076	0.051
$\overline{F}$	0.735	0.709	0.738	0.726
$\overline{\Delta F}$	0.202	0.186	0.212	0.170
$B^{sp}_{2009} / B^{target}$	1.000	1.000	1.013	1.047
$\min B^{sp}_{y} / B^{target}$	0.491	0.610	0.429	0.451

Table 1a. Comparison of North Sea Plaice results for deterministic "hindsight" projections under a Beverton–Holt stock–recruit relationship when using only the BTS-Isis aggregated index in "hindsight" MPs (see text for details of the MP control parameters). Units are tons where applicable.

PLAICE: SUBAREA IV		2-LINE STOCK RECRUIT RELATIONSHIP		
		MP1	MP2	MP3
	Observed catches	Constant catch	MP slope: BTS-Isis $x = 80\%, \lambda = 0.562$ $p = 4$ (max 20% ΔTAC)	MP target: BTS-Isis $I^0 = 0.2I^{ave}, I^{target} = 1.1I^{ave}$ $TAC^{target} = 206700$ $w = 0.49$ (max 20% ΔTAC)
$\overline{TAC}$	145702	178293	151582	154935
$\overline{\Delta TAC}$	0.166	0.018	0.075	0.051
$\overline{F}$	0.735	0.634	0.737	0.722
$\overline{\Delta F}$	0.202	0.198	0.217	0.172
$B_{2009}^{sp} / B^{target}$	1.000	1.000	1.023	0.997
$\min B_y^{sp} / B^{target}$	0.491	0.680	0.422	0.445

Table 1b. As for Table 1a, but here when projecting with a 2-line stock–recruit relationship.

PLAICE: SUBAREA IV		BEVERTON–HOLT STOCK RECRUIT RELATIONSHIP WITH H=0.9		
	Observed catches	Constant catch $TAC^{target} = 185989$	MP slope: BTS-Isis $x = 80\%, \lambda = 0.556$ $p = 4$ (max 20% ΔTAC)	MP target: BTS-Isis $I^0 = 0.2I^{ave}$ $I^{target} = 1.1I^{ave}$ $TAC^{target} = 206700$ $w = 0.5$ (max 20% ΔTAC)
$\overline{TAC}$	145702	185989 (185989, 185989)	174343 (107374, 214871)	169445 (91831, 218102)
$\overline{\Delta TAC}$	0.166	0.016 (0.016, 0.016)	0.078 (0.057, 0.118)	0.067 (0.048, 0.143)
$\overline{F}$	0.735	3.146 (0.197, 7.270)	0.317 (0.225, 3.220)	0.351 (0.227, 6.398)
$\overline{\Delta F}$	0.202	0.283 (0.142, 0.471)	0.186 (0.126, 0.341)	0.196 (0.124, 0.442)
$B_{2009}^{sp} / B^{target}$	1.000	0.000 (0.000, 12.315)	5.468 (0.000, 10.792)	4.604 (0.000, 10.134)
$\min B_y^{sp} / B^{target}$	0.491	0.000 (0.000, 1.013)	0.853 (0.000, 0.950)	0.798 (0.000, 0.955)

Table 2a. Comparison of results for stochastic "forecast" projections for North Sea Plaice under a Beverton–Holt stock–recruit relationship when using only the BTS-Isis aggregated index in the slope and target type "hindsight" MPs. Management quantities shown are medians with associated 95% probability intervals in parentheses. 1000 simulations were performed. Units are tons where applicable.

PLAICE: SUBAREA IV		2-LINE STOCK RECRUIT RELATIONSHIP		
	Observed catches	Constant catch $TAC^{target} = 178293$	MP slope: BTS-Isis $x = 80\%, \lambda = 0.562$ $p = 4$ (max 20% ΔTAC)	MP target: BTS-Isis $I^0 = 0.2 I^{ave}$ $I^{target} = 1.1 I^{ave}$ $TAC^{target} = 206700$ $w = 0.49$ (max 20% ΔTAC)
$\overline{TAC}$	145702	178293 (178293, 178293)	149568 (96345, 172055)	149868 (87507, 172475)
$\overline{\Delta TAC}$	0.166	0.018 (0.018, 0.018)	0.078 (0.057, 0.132)	0.062 (0.047, 0.152)
$\overline{F}$	0.735	4.529 (0.267, 7.220)	0.354 (0.258, 4.056)	0.419 (0.270, 6.555)
$\overline{\Delta F}$	0.202	0.332 (0.138, 0.502)	0.186 (0.126, 0.341)	0.197 (0.119, 0.434)
$B^{sp}_{2009} / B^{target}$	1.000	0.000 (0.000, 4.102)	2.695 (0.000, 4.890)	2.050 (0.000, 4.529)
$\min B^{sp}_{y} / B^{target}$	0.491	0.000 (0.000, 1.013)	0.798 (0.000, 0.949)	0.698 (0.000, 0.953)

Table 2b. As for Table 2a, but here when projecting with a 2-line stock–recruit relationship.

PLAICE: SUBAREA IV		BEVERTON–HOLT STOCK RECRUIT RELATIONSHIP WITH H=0.9		
	Observed catches	Constant catch $TAC^{target} = 160634$ Median constant catch required to reach target for each simulation: 184068 (160634,216004)	MP slope: BTS-Isis $x = 80\%, \lambda = 1.0$ $p = 4$ (max 20% ΔTAC)	MP target: BTS-Isis $I^0 = 0.2I^{ave}$ $I^{target} = 1.1I^{ave}$ $TAC^{target} = 206700$ $w = 0.0$ (max 20% ΔTAC)
$\overline{TAC}$	145702	160634 (160634, 160634)	181650 (137235, 234484)	184272 (126068, 248620)
$\overline{\Delta TAC}$	0.166	0.021 (0.021, 0.021)	0.118 (0.091, 0.148)	0.146 (0.101, 0.191)
$\overline{F}$	0.735	0.200 (0.136, 0.457)	0.238 (0.190, 0.338)	0.214 (0.182, 0.263)
$\overline{\Delta F}$	0.202	0.189 (0.142, 0.295)	0.189 (0.137, 0.260)	0.198 (0.147, 0.265)
$B_{2009}^{sp} / B^{target}$	1.000	9.647 (1.000, 16.739)	7.624 (3.794, 12.943)	8.684 (6.293, 12.808)
$\min B_y^{sp} / B^{target}$	0.491	1.013 (0.736, 1.013)	0.902 (0.739, 0.945)	0.893 (0.682, 0.955)

Table 3a. Comparison of results for stochastic "forecast" projections for North Sea Plaice under a Beverton–Holt stock–recruit relationship for the "forecast" MPs. Management quantities shown are medians with associated 95% probability intervals in parentheses. 1000 simulations were performed. Units are tons where applicable.

PLAICE: SUBAREA IV		2-LINE STOCK RECRUIT RELATIONSHIP		
	Observed catches	Constant catch $TAC^{target} = 146275$ Median constant catch required to reach target for each simulation: 168306 (146275,195181)	MP slope: BTS-Isis $x = 80\%, \lambda = 0.61$ $p = 4$ (max 20% ΔTAC)	MP target: BTS-Isis $I^0 = 0.2I^{ave}$ $I^{target} = 1.1I^{ave}$ $TAC^{target} = 206700$ $w = 0.378$ (max 20% ΔTAC)
$\overline{TAC}$	145702	146275 (146275, 146275)	147736 (122006, 171214)	144947 (120021, 169205)
$\overline{\Delta TAC}$	0.166	0.024 (0.024, 0.024)	0.083 (0.060, 0.113)	0.076 (0.057, 0.105)
$\overline{F}$	0.735	0.216 (0.152, 0.398)	0.322 (0.245, 0.540)	0.316 (0.239, 0.554)
$\overline{\Delta F}$	0.202	0.172 (0.122, 0.248)	0.179 (0.122, 0.275)	0.179 (0.123, 0.282)
$B_{2009}^{sp} / B^{target}$	1.000	4.126 (1.000, 7.238)	3.085 (1.057, 5.261)	3.337 (1.002, 5.362)
$\min B_y^{sp} / B^{target}$	0.491	1.013 (0.804, 1.013)	0.831 (0.435, 0.949)	0.794 (0.405, 0.954)

Table 3b. As for Table 3a, but here projecting with a 2-line stock–recruit relationship.

PLAICE: SUBAREA IV		BEVERTON–HOLT STOCK RECRUIT RELATIONSHIP WITH H=0.9		
	Observed catches	Constant catch $TAC^{target} = 1606$	MP slope: BTS-Isis $x = 80\%, \lambda = 1.0$ $p = 4$ (max 20% ΔTAC)	MP target: BTS-Isis $I^0 = 0.2I^{ave}, I^{target} = 1.$ $TAC^{target} = 206700$ $w = 0.0$ (max 20% ΔTAC)
$\overline{TAC}$	145702	160634	173748	198740
$\overline{\Delta TAC}$	0.166	0.021	0.120	0.112
$\overline{F}$	0.735	0.235	0.315	0.331
$\overline{\Delta F}$	0.202	0.188	0.207	0.166
$B^{sp}_{2009} / B^{target}$	1.000	10.454	7.782	6.375
$\min B^{sp}_{y} / B^{target}$	0.491	0.921	0.813	0.796

Table 4a. Comparison of results for deterministic "hindsight" projections under a Beverton–Holt stock–recruit relationship when using only the BTS-Isis aggregated index in the "forecast" MPs. Units are tons where applicable.

PLAICE: SUBAREA IV		2-LINE STOCK RECRUIT RELATIONSHIP		
	Observed catches	Constant catch $TAC^{target} = 1462$	MP slope: BTS-Isis $x = 80\%, \lambda = 0.$ $p = 4$ (max 20% ΔTAC)	MP target: BTS-Isis $I^0 = 0.2I^{ave}, I^{target} = 1.$ $TAC^{target} = 206700$ $w = 0.378$ (max 20% ΔTAC)
$\overline{TAC}$	145702	146275	146550	147666
$\overline{\Delta TAC}$	0.166	0.024	0.089	0.067
$\overline{F}$	0.735	0.306	0.727	0.537
$\overline{\Delta F}$	0.202	0.194	0.208	0.163
$B^{sp}_{2009} / B^{target}$	1.000	2.811	1.413	1.788
$\min B^{sp}_y / B^{targ}$	0.491	0.921	0.446	0.529

Table 4b. As for Table4a, but here when projecting with a 2-line stock–recruit relationship.

SOLE: SUBAREA IV		BEVERTON–HOLT STOCK RECRUIT RELATIONSHIP WITH H=0.9		
	Observed catches	Constant catch $TAC^{target} = 24312$	MP slope: BTS-Isis $x = 110\%, \lambda = 0.21$ $p = 4$ (max 20% ΔTAC)	MP target: BTS-Isis $I^0 = 0.2 I^{ave}, I^{target} = 2 I^{ave}$ $TAC^{target} = 26300$ $w = 0.5$ (max 20% ΔTAC)
$\overline{TAC}$	22364	24312	23981	23221
$\overline{\Delta TAC}$	0.152	0.005	0.055	0.063
$\overline{F}$	0.908	0.602	0.700	0.799
$\overline{\Delta F}$	0.274	0.302	0.309	0.289
$B_{2009}^{sp} / B^{target}$	1.000	1.000	0.997	0.985
$\min B_y^{sp} / B^{target}$	0.519	0.925	0.718	0.541

Table 5a. Comparison of results for deterministic "hindsight" projections for Sole in Subarea IV under a Beverton–Holt stock–recruit relationship when using only the BTS-Isis aggregated index in the MPs selected with hindsight (see the text for details of the MP control parameters) when using a Beverton–Holt stock–recruit relationship. Units are tons where applicable.

SOLE: SUBAREA IV **2-LINE STOCK RECRUIT RELATIONSHIP**

	Observed catches	**Constant catch** $TAC^{target} = 22465$	**MP slope: BTS-Isis** $x = 110\%, \lambda = 0.19$ $p = 4$ (max 20% ΔTAC)	**MP target: BTS-Isis** $I^0 = 0.2I^{ave}, I^{target} = 2I^{ave}$ $TAC^{target} = 25600$ $w = 0.5$ (max 20% ΔTAC)
$\overline{TAC}$	22364	22465	22700	22504
$\overline{\Delta TAC}$	0.152	0.001	0.045	0.062
$\overline{F}$	0.908	0.525	0.615	0.749
$\overline{\Delta F}$	0.274	0.291	0.306	0.285
$B^{sp}_{2009} / B^{target}$	1.000	1.001	0.991	1.015
$\min B^{sp}_y / B^{target}$	0.519	0.982	0.773	0.571

Table 5b. As for Table 5a, but here projecting with a 2-line stock–recruit relationship.

SOLE: SUBAREA IV **BEVERTON–HOLT STOCK RECRUIT RELATIONSHIP WITH H=0.9**

	Observed catches	Constant catch $TAC^{target} = 24312$	MP slope: BTS-Isis $x = 110\%, \lambda = 0.21$ $p = 4$ (max 20% ΔTAC)	MP target: BTS-Isis $I^0 = 0.2I^{ave}$ $I^{target} = 2I^{ave}$ $TAC^{target} = 26300$ $w = 0.5$ (max 20% ΔTAC)
$\overline{TAC}$	22364	24312 (24312, 24312)	23147 (19730, 24836)	21883 (19551, 24554)
$\overline{\Delta TAC}$	0.152	0.005 (0.005, 0.005)	0.042 (0.030, 0.072)	0.044 (0.033, 0.058)
$\overline{F}$	0.908	0.402 (0.219, 3.989)	0.705 (0.301, 4.477)	0.387 (0.283, 1.672)
$\overline{\Delta F}$	0.274	0.177 (0.114, 0.454)	0.226 (0.117, 0.442)	0.156 (0.111, 0.328)
$B_{2009}^{sp} / B^{target}$	1.000	1.448 (0.000, 4.734)	0.432 (0.001, 3.162)	1.654 (0.020, 3.617)
$\min B_y^{sp} / B^{target}$	0.519	1.314 (0.000, 2.496)	0.398 (0.000, 2.345)	1.400 (0.007, 2.456)

Table 6a. Comparison of results for stochastic "forecast" projections for Sole in Subarea IV under a Beverton–Holt stock–recruit relationship when using only the BTS-Isis aggregated index in the slope and target type "hindsight" MPs. Management quantities shown are medians with associated 95% probability intervals in parentheses. 1000 simulations were performed. Units are tons where applicable.

SOLE: SUBAREA IV **2-LINE STOCK RECRUIT RELATIONSHIP**

	Observed catches	Constant catch $TAC^{target} = 22465$	MP slope: BTS-Isis $x = 110\%, \lambda = 0.19$ $p = 4$ (max 20% ΔTAC)	MP target: BTS-Isis $I^0 = 0.2I^{ave}$ $I^{target} = 2I^{ave}$ $TAC^{target} = 25600$ $w = 0.5$ (max 20% ΔTAC)
$\overline{TAC}$	22364	22465 (22465, 22465)	22912 (18954, 24550)	21007 (19120, 23332)
$\overline{\Delta TAC}$	0.152	0.001 (0.001, 0.001)	0.043 (0.029, 0.083)	0.043 (0.032, 0.058)
$\overline{F}$	0.908	0.514 (0.246, 3.926)	2.013 (0.361, 4.928)	0.379 (0.281, 0.928)
$\overline{\Delta F}$	0.274	0.210 (0.123, 0.464)	0.308 (0.131, 0.457)	0.156 (0.113, 0.268)
$B^{sp}_{2009} / B^{target}$	1.000	0.745 (0.000, 3.586)	0.012 (0.000, 2.265)	1.612 (0.161, 2.351)
$\min B^{sp}_{y} / B^{target}$	0.519	0.700 (0.000, 2.456)	0.006 (0.000, 1.832)	1.364 (0.161, 2.389)

Table 6b. As for Table 6a, but here projecting with a 2-line stock–recruit relationship.

SOLE: SUBAREA IV **BEVERTON–HOLT STOCK RECRUIT RELATIONSHIP WITH H=0.9**

	Observed catches	**Constant catch** $TAC^{target} = 19570$ Median constant catch required to reach target for each simulation: 23136 (19570, 27311)	**MP slope: BTS-Isis** $x = 110\%, \lambda = 0.45$ $p = 3$ (max 20% ΔTAC)	**MP target: BTS-Isis** $I^0 = 0.2 I^{ave}$ $I^{target} = 1.5 I^{ave}$ $TAC^{target} = 24000$ $w = 0.5$ (max 20% ΔTAC)
$\overline{TAC}$	22364	19570 (19570, 19570)	20942 (18706, 23266)	20747 (18755, 23163)
$\overline{\Delta TAC}$	0.152	0.006 (0.006, 0.006)	0.069 (0.051, 0.093)	0.045 (0.035, 0.061)
$\overline{F}$	0.908	0.219 (0.160, 0.375)	0.318 (0.249, 0.435)	0.284 (0.228, 0.385)
$\overline{\Delta F}$	0.274	0.149 (0.110, 0.204)	0.144 (0.105, 0.195)	0.146 (0.109, 0.190)
$B^{sp}_{2009} / B^{target}$	1.000	3.369 (1.000, 6.342)	2.331 (0.988, 4.359)	2.611 (1.094, 4.576)
$\min B^{sp}_{y} / B^{target}$	0.519	2.456 (0.967, 2.456)	1.921 (0.940, 2.456)	2.150 (1.033, 2.456)

Table 7a. Comparison of results for stochastic "forecast" projections for Sole in Subarea IV under a Beverton–Holt stock–recruit relationship for the best performing "forecast" MPs when incorporating only the BTS-Isis aggregated index in the HCR with a Beverton–Holt stock–recruit relationship. Management quantities shown are medians with associated 95% probability intervals in parentheses. 1000 simulations were performed. Units are tons where applicable.

SOLE: SUBAREA IV **2-LINE STOCK RECRUIT RELATIONSHIP**

	Observed catches	**Constant catch** $TAC^{target} = 18743$ Median constant catch required to reach target for each simulation: 22118 (18743,26168)	**MP slope: BTS-Isis** $x = 110\%, \lambda = 0.48$ $p = 3$ (max 20% ΔTAC)	**MP target: BTS-Isis** $I^0 = 0.2 I^{ave}$ $I^{target} = 1.5 I^{ave}$ $TAC^{target} = 23500$ $w = 0.5$ (max 20% ΔTAC)
$\overline{TAC}$	22364	18743 (18743, 18743)	20104 (18109, 22277)	19877 (18177, 21978)
$\overline{\Delta TAC}$	0.152	0.008 (0.008, 0.008)	0.074 (0.054, 0.098)	0.047 (0.036, 0.061)
$\overline{F}$	0.908	0.220 (0.161, 0.367)	0.322 (0.256, 0.423)	0.286 (0.230, 0.383)
$\overline{\Delta F}$	0.274	0.149 (0.110, 0.204)	0.156 (0.106, 0.196)	0.147 (0.110, 0.191)
$B_{2009}^{sp} / B^{target}$	1.000	2.948 (1.000, 5.519)	2.117 (1.002, 3.867)	2.364 (1.067, 4.094)
$\min B_y^{sp} / B^{target}$	0.519	2.456 (0.929, 2.456)	1.778 (0.944, 2.456)	1.980 (1.007, 2.456)

Table 7b. As for Table 7a, but here for a 2-line stock recruit relationship.

SOLE: SUBAREA IV **BEVERTON–HOLT STOCK RECRUIT RELATIONSHIP WITH H=0.9**

	Observed catches	Constant catch $TAC^{target} = 19570$	MP slope: BTS-Isis $x = 110\%, \lambda = 0.45$ $p = 3$ (max 20% ΔTAC)	MP target: BTS-Isis $I^0 = 0.2I^{ave}, I^{target} = 1.5I^{ave}$ $TAC^{target} = 24000$ $w = 0.5$ (max 20% ΔTAC)
$\overline{TAC}$	22364	19570	21681	22219
$\overline{\Delta TAC}$	0.152	0.006	0.100	0.067
$\overline{F}$	0.908	0.262	0.487	0.501
$\overline{\Delta F}$	0.274	0.244	0.297	0.272
$B_{2009}^{sp} / B^{target}$	1.000	3.178	2.240	1.920
$\min B_y^{sp} / B^{target}$	0.519	2.596	1.390	1.284

Table 8a. Comparison of results for deterministic "hindsight" projections for Sole in Subarea IV under a Beverton–Holt stock–recruit relationship when using only the BTS-Isis aggregated index in the "forecast" MPs. Units are tons where applicable.

SOLE: SUBAREA IV **2-LINE STOCK RECRUIT RELATIONSHIP**

	Observed catches	Constant catch $TAC^{target} = 18743$	MP slope: BTS-Isis $x = 110\%, \lambda = 0.48$ $p = 3$ (max 20% ΔTAC)	MP target: BTS-Isis $I^0 = 0.2I^{ave}, I^{target} = 1.5I^{ave}$ $TAC^{target} = 23500$ $w = 0.5$ (max 20% ΔTAC)
$\overline{TAC}$	22364	18743	20859	21356
$\overline{\Delta TAC}$	0.152	0.008	0.105	0.066
$\overline{F}$	0.908	0.269	0.501	0.508
$\overline{\Delta F}$	0.274	0.241	0.301	0.272
$B^{sp}_{2009} / B^{target}$	1.000	2.444	1.897	1.592
$\min B^{sp}_y / B^{target}$	0.519	2.191	1.323	1.082

Table 8b. As for Table 8a, but here projecting with a 2-line stock–recruit relationship.

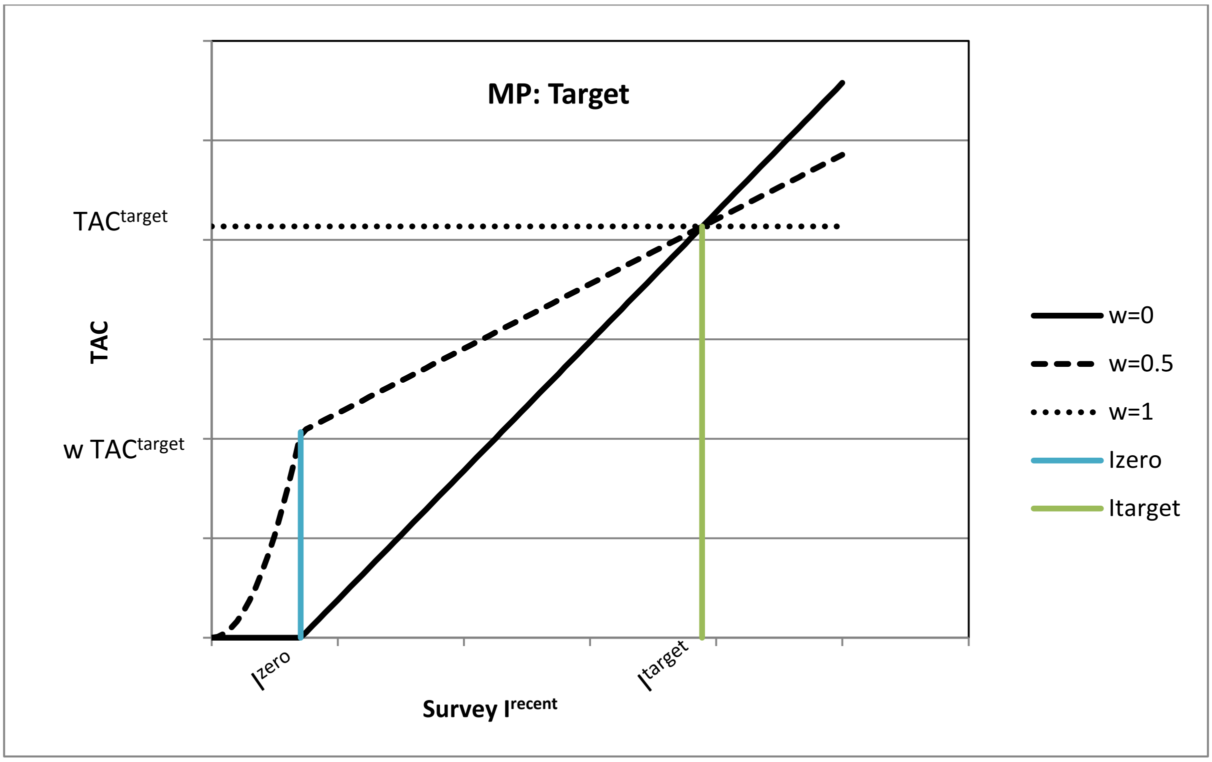


Figure 1. Different target type MPs for three values of *w*: the dashed lines correspond to equations 13, 14 and 16, while the solid black line correspond to equation 15. The vertical lines indicate the zero and target survey values, while the horizontal dotted line corresponds to the target TAC (constant catch rule for *w*=1)

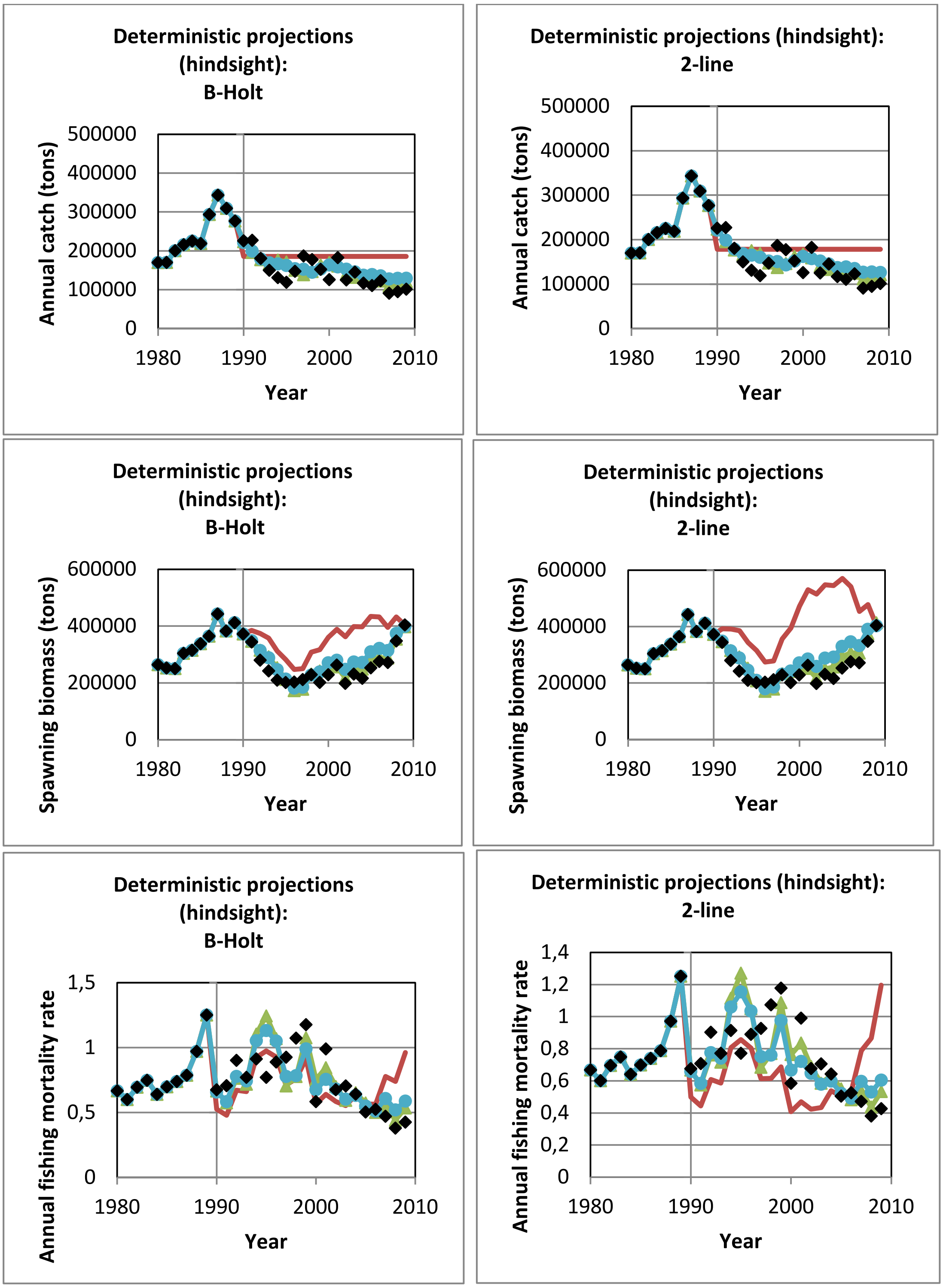
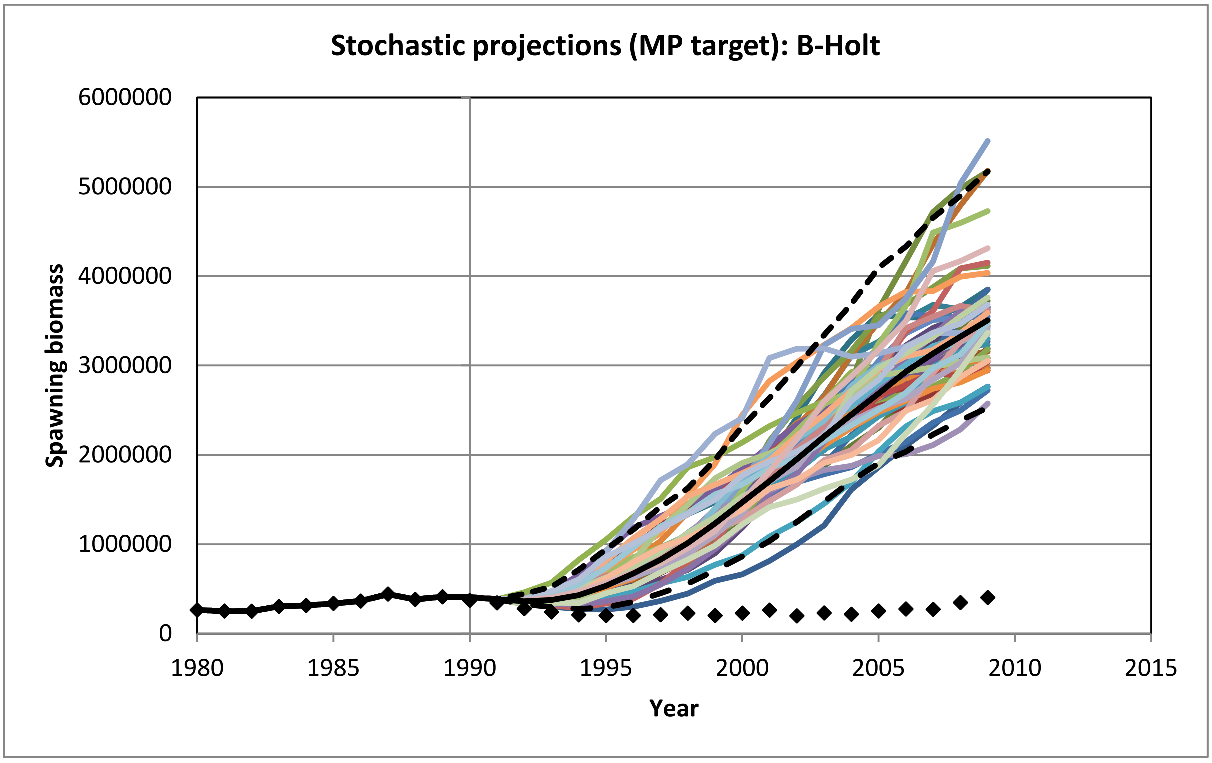


Figure 2:.Deterministic "hindsight" projections with either a Beverton–Holt or 2-line stock–recruit relationship from 1990 for a constant catch strategy (line), and for BTS-Isis survey slope (triangles) and target (dots) "hindsight" MPs, compared to the adjusted 2010 XSA assessment estimates for North Sea Plaice (black diamonds). Top plots: total annual catch; middle plots: spawning biomass; bottom plots: annual fishing mortality.



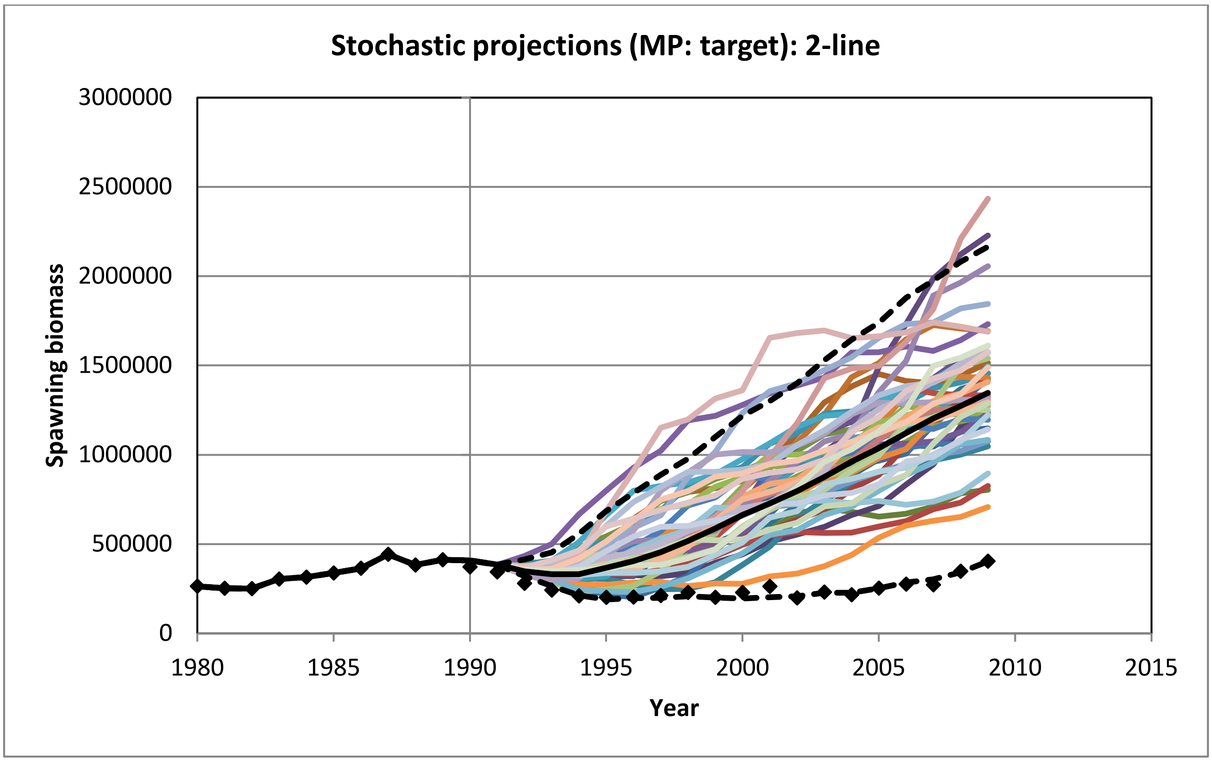
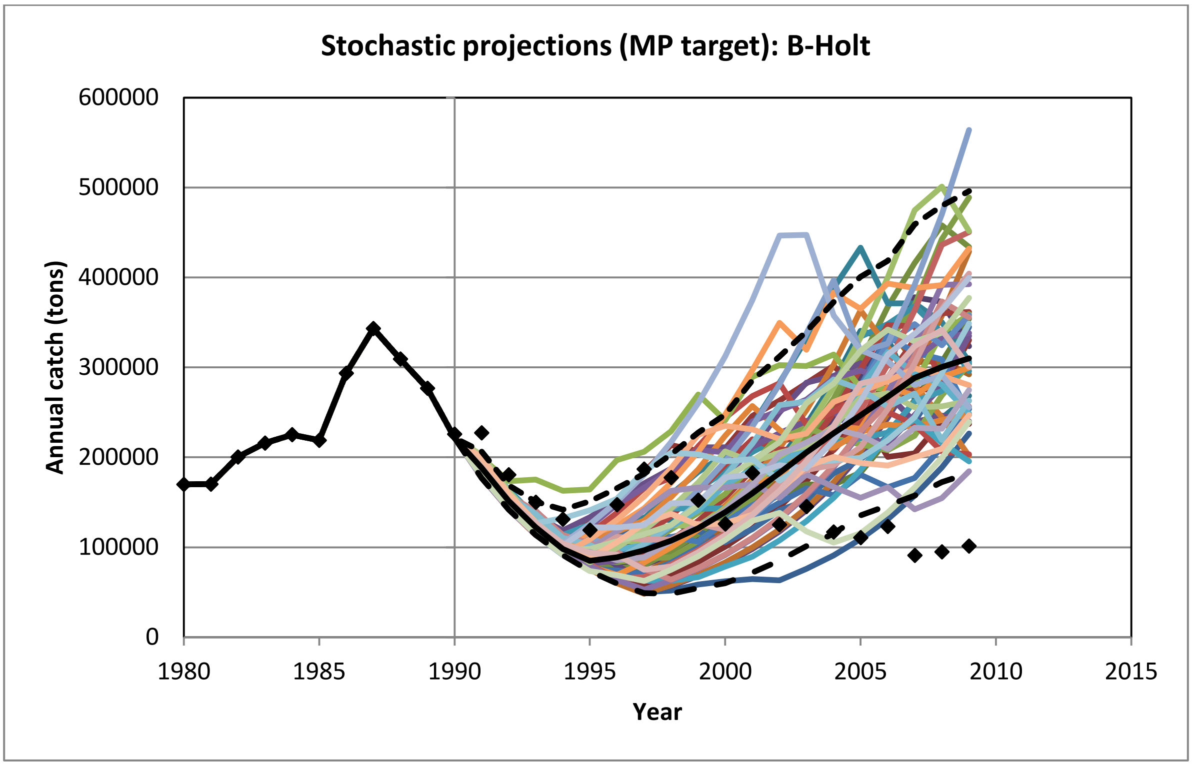


Figure 3. Stochastic "forecast" spawning biomass projections from 1990 for a BTS-Isis survey target "forecast" MP with a Beverton–Holt (top) and a 2-line (bottom) stock–recruit relationship (50 of the 1000 simulations shown here) compared to the adjusted XSA assessment estimates for North Sea Plaice (black diamonds). The medians and 95% PIs are indicated by the solid and dashed black lines.



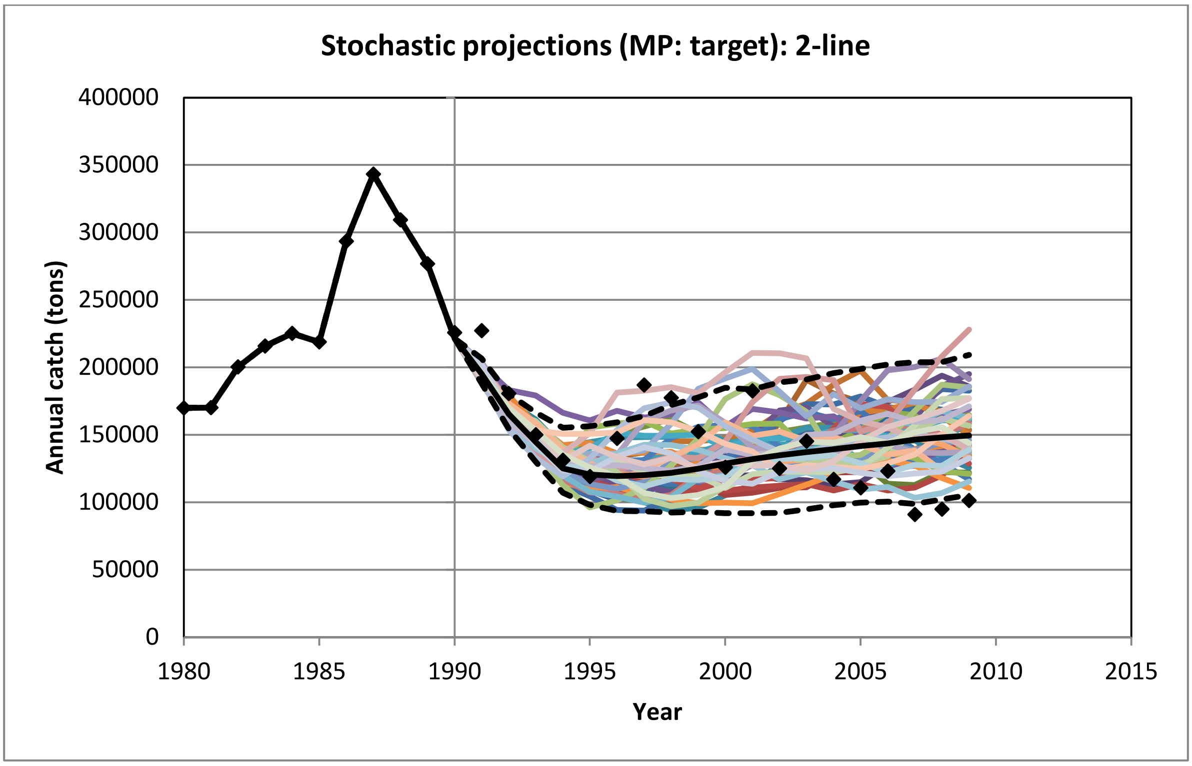


Figure 4. Stochastic "forecast" catch projections from 1990 for a BTS-Isis survey target "forecast" MP with a Beverton–Holt (top) and 2-line (bottom) stock–recruit relationship (50 of the 1000 simulations shown here) compared to the observed catches (landings plus discards) for North Sea Plaice (black diamonds). The medians and 95% PIs are indicated by the solid and dashed black lines.

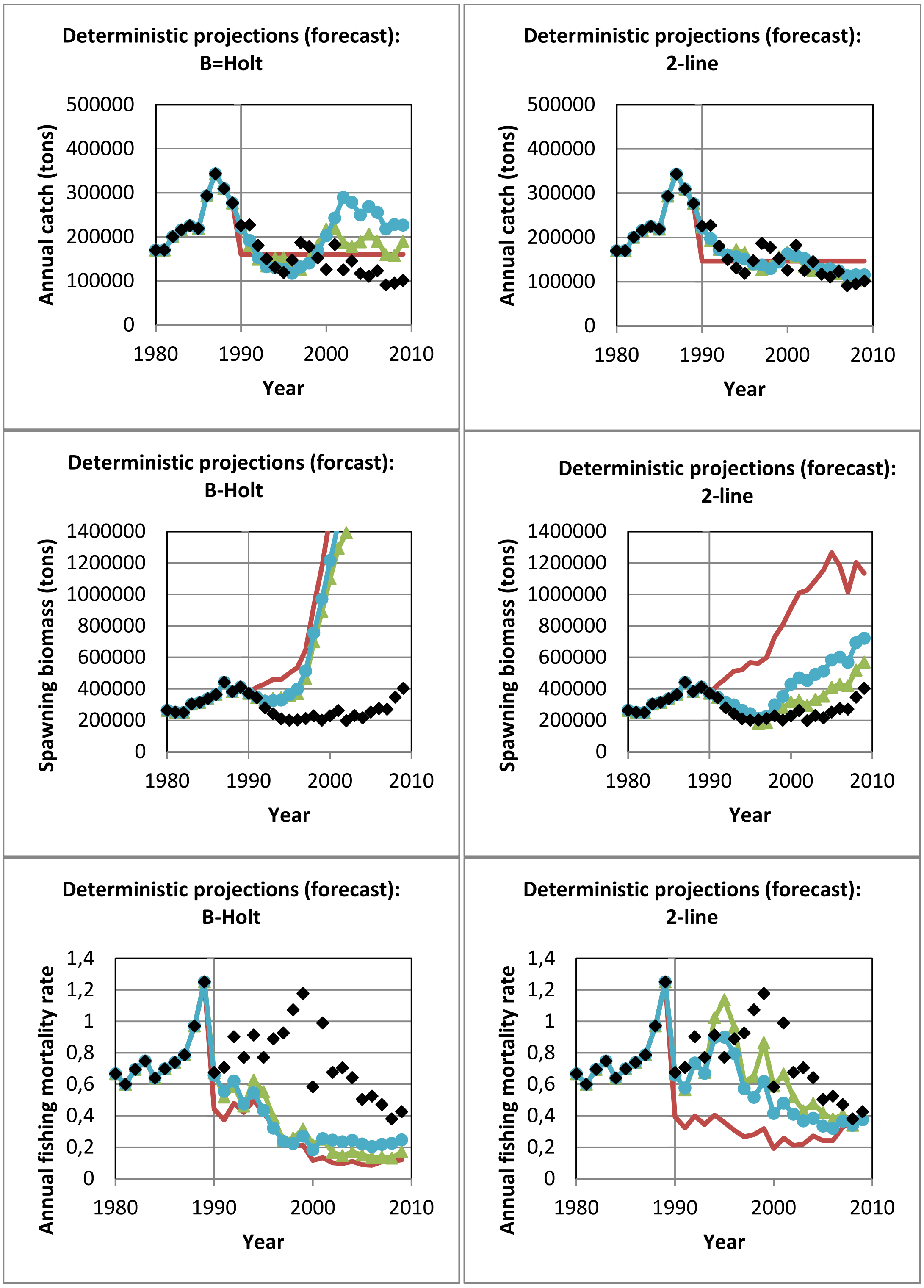


Figure 5. Deterministic "hindsight" projections with a Beverton–Holt (left) and 2-line (right) stock–recruit relationship from 1990 for a constant catch strategy (line), and for BTS-Isis survey slope (triangles) and target (dots) "forecast" MPs, compared to the adjusted 2010 XSA assessment estimates for North Sea Plaice (black diamonds). Top two plots: total annual catch. Middle two plots: spawning biomass. Bottom two plots: annual fishing mortality.

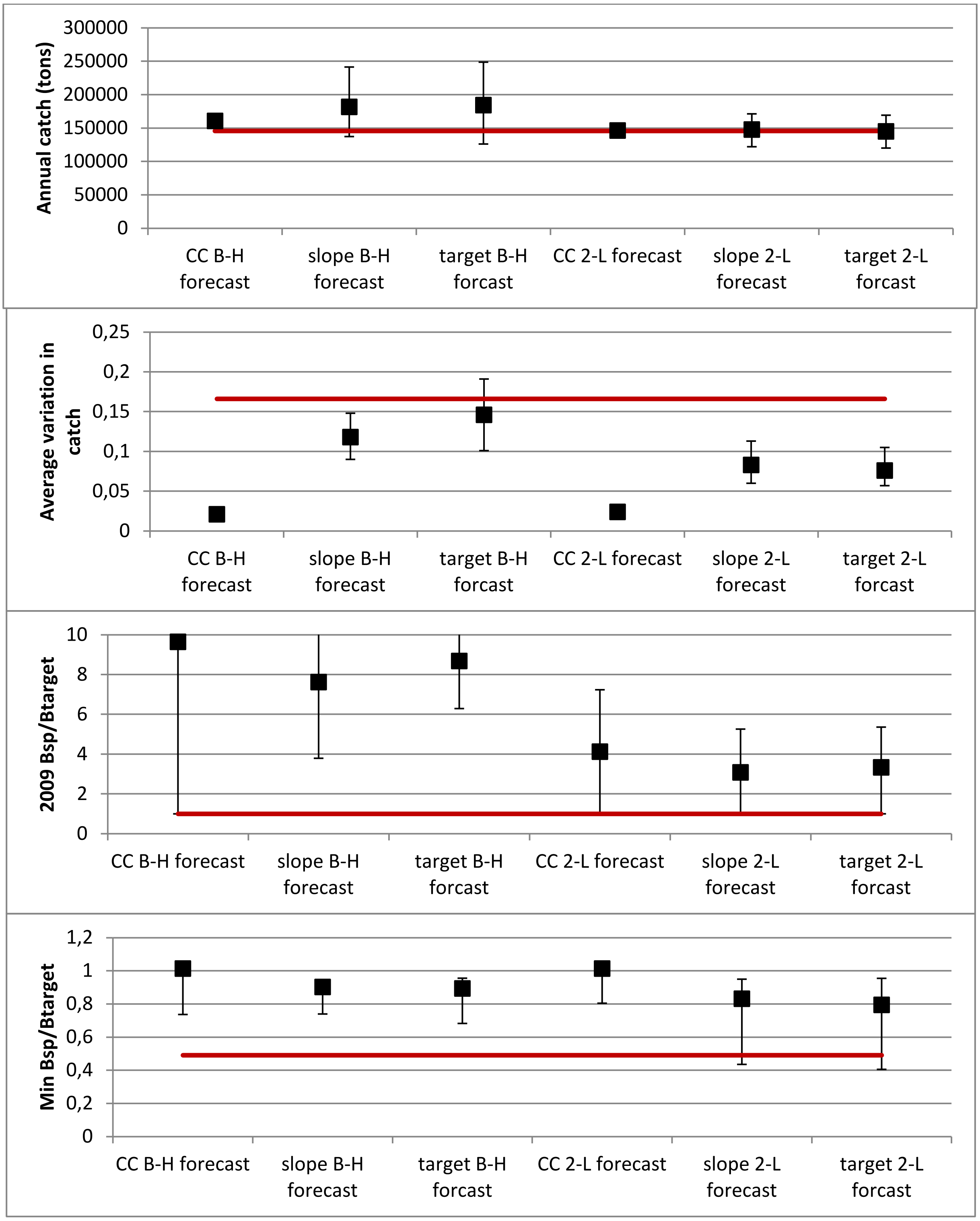


Figure 6a. Comparison of performance statistics for "forecast" MPs for North Sea Plaice over the 1990 to 2009 period when projecting with a Beverton–Holt and 2-line stock–recruit relationship respectively: medians and 95% probability intervals of 1000 simulations. From top to bottom: average annual future catch, average interannual variation in catch, final spawning biomass as a fraction of target, and minimum future spawning biomass as a fraction of the target value. The solid horizontal lines indicate results for the adjusted 2010 XSA assessment estimates.

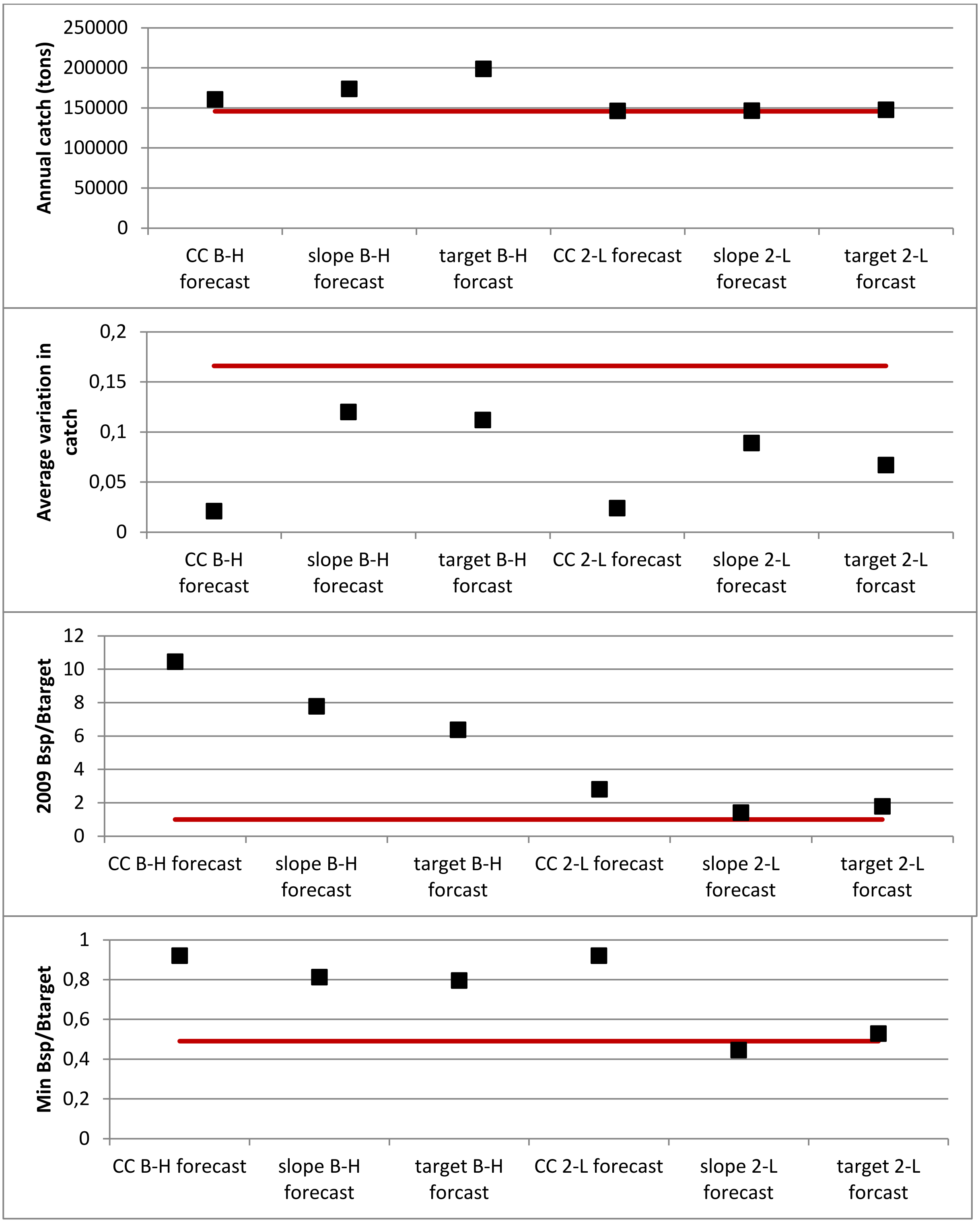


Figure 6b. Comparison of performance statistics for deterministic "hindsight" projections under "forecast" MPs for North Sea Plaice over the 1990 to 2009 period with a Beverton–Holt and 2-line stock–recruit relationship respectively. From top to bottom: average annual future catch, average interannual variation in catch, final spawning biomass as a fraction of target, and minimum future spawning biomass as a fraction of the target value. The solid horizontal lines indicate results for the adjusted 2010 XSA assessment estimates.

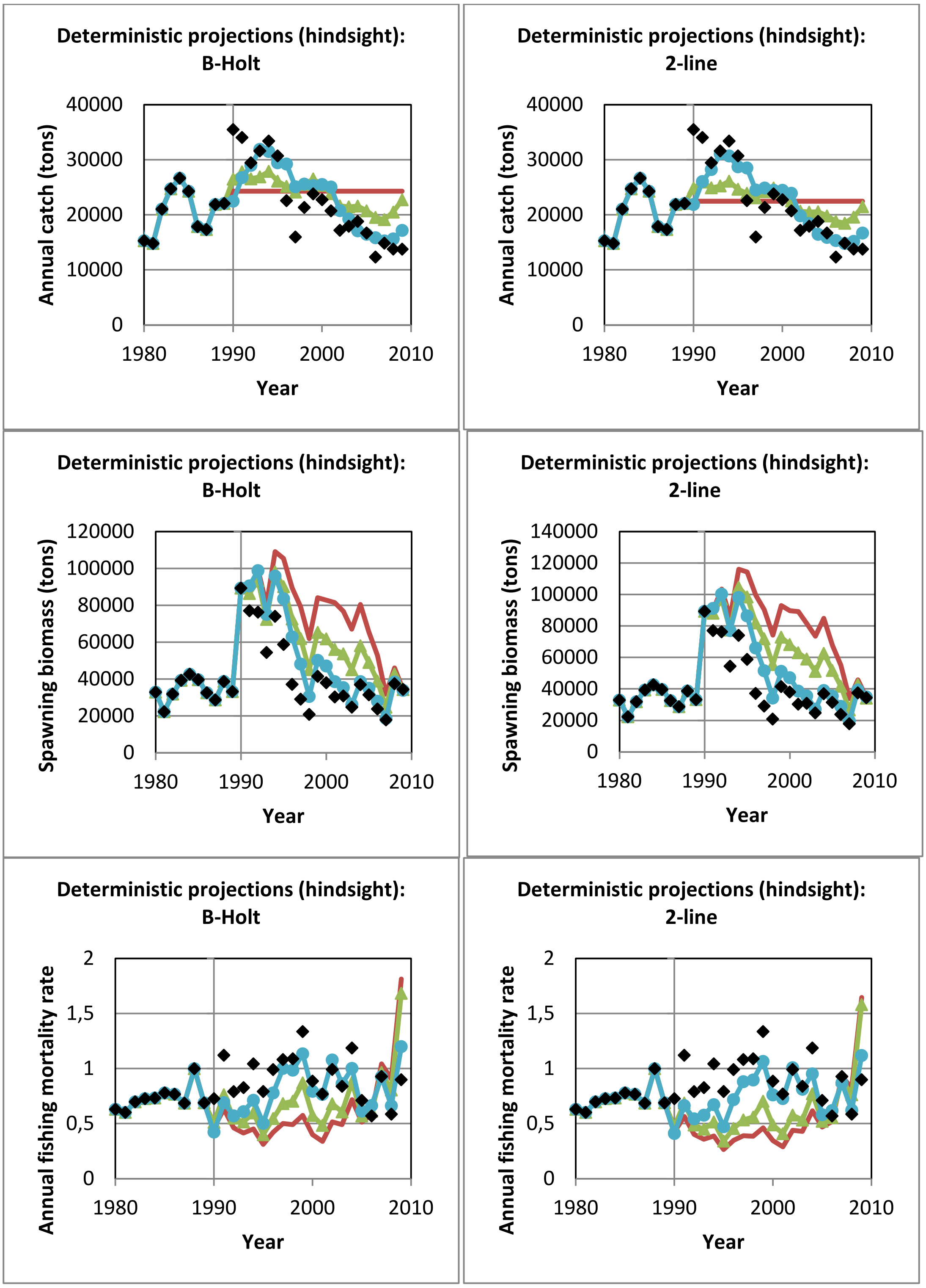
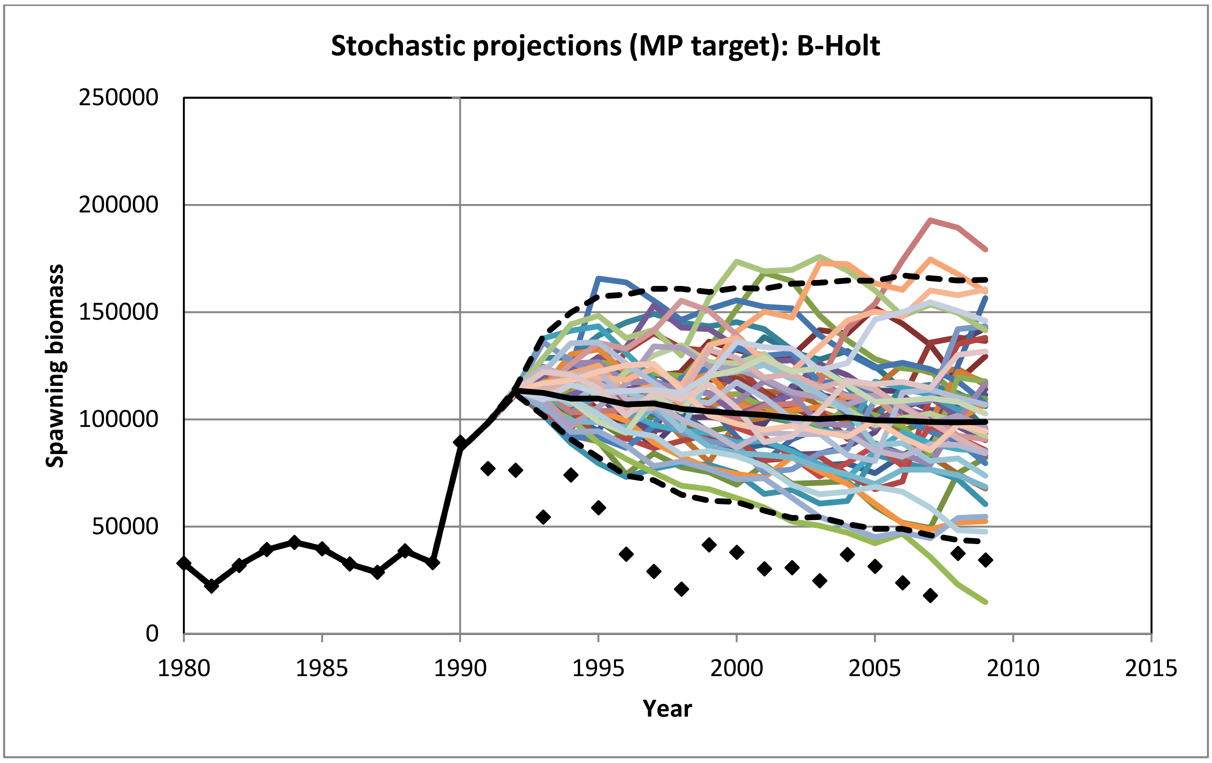


Figure 7. Deterministic "hindsight" projections with either a Beverton–Holt or 2-line stock–recruit relationship from 1990 for a constant catch strategy (line), and for BTS-Isis survey slope (triangles) and target (dots) "hindsight" MPs, compared to the adjusted 2010 XSA assessment estimates for Sole caught in Subarea IV (black diamonds). Top plots: total annual catch; middle plots: spawning biomass; bottom plots: annual fishing mortality.



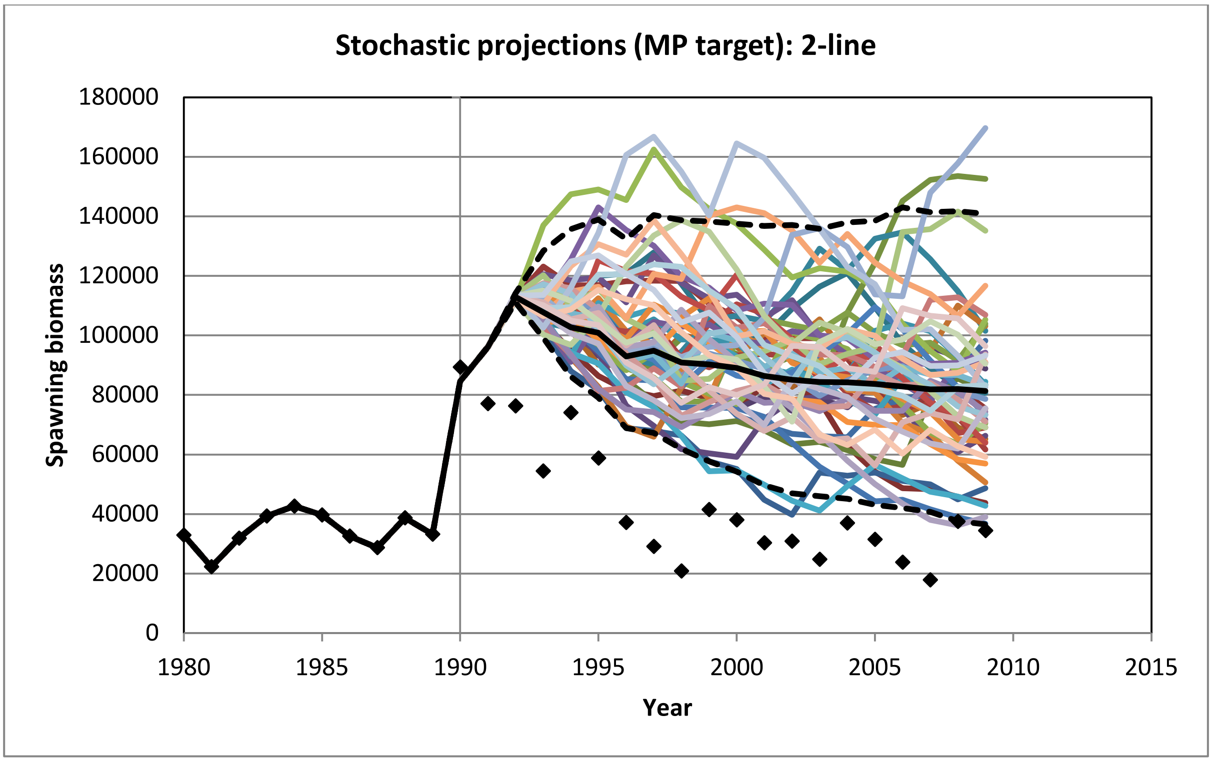
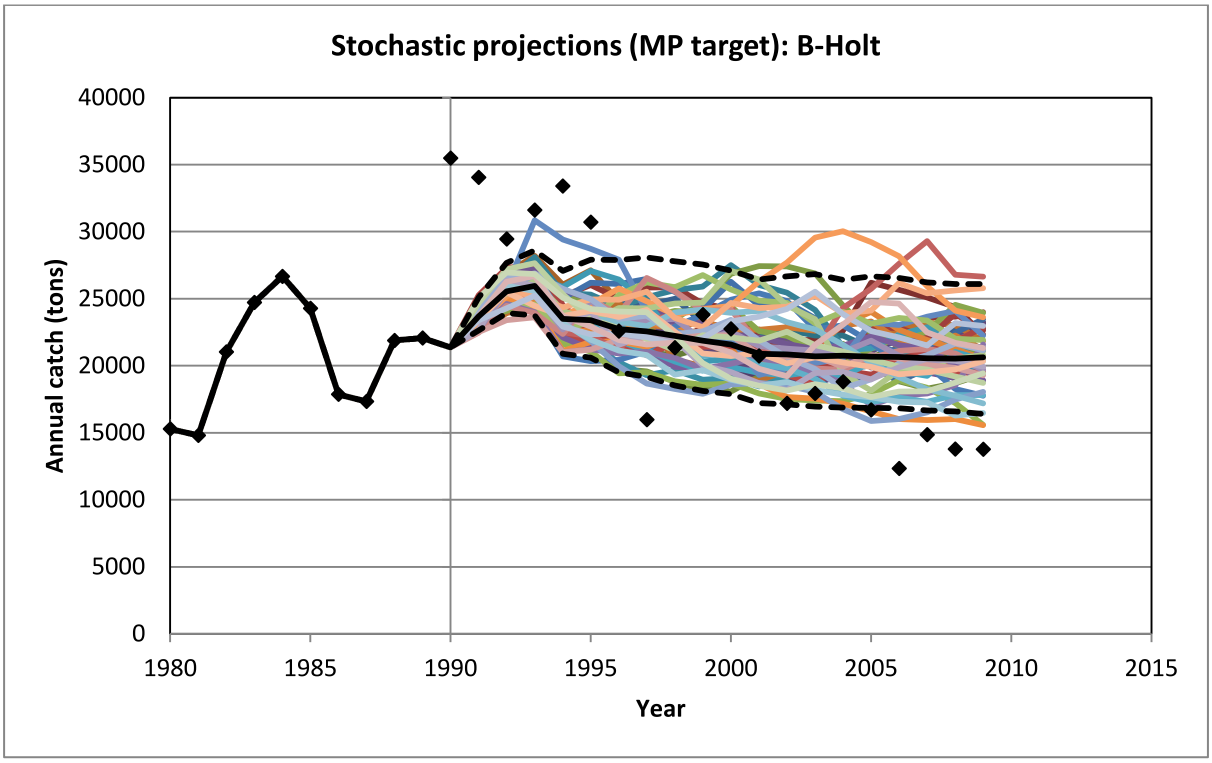


Figure 8. Stochastic "forecast" spawning biomass projections from 1990 for a BTS-Isis survey target "forecast" MP with a Beverton–Holt (top) and a 2-line (bottom) stock–recruit relationship (50 of the 1000 simulations shown here) compared to the adjusted XSA assessment estimates for Sole in Subarea IV (black diamonds). The medians and 95% PIs are indicated by the solid and dashed black lines.



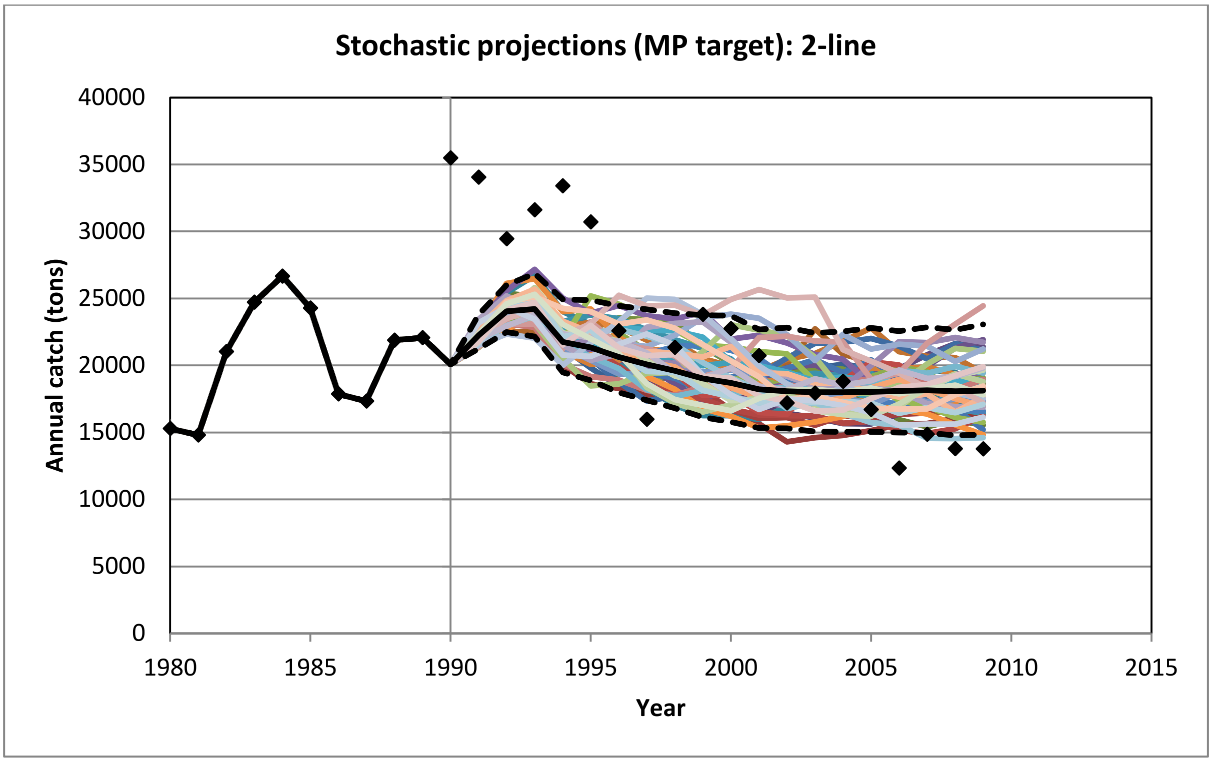


Figure 9. Stochastic "forecast" catch projections from 1990 for a BTS-Isis survey target "forecast" MP with a Beverton–Holt (top) and 2-line (bottom) stock–recruit relationship (50 of the 1000 simulations shown here) compared to the observed catches (landings plus discards) for Sole in Subarea IV (black diamonds). The medians and 95% PIs are indicated by the solid and dashed black lines.

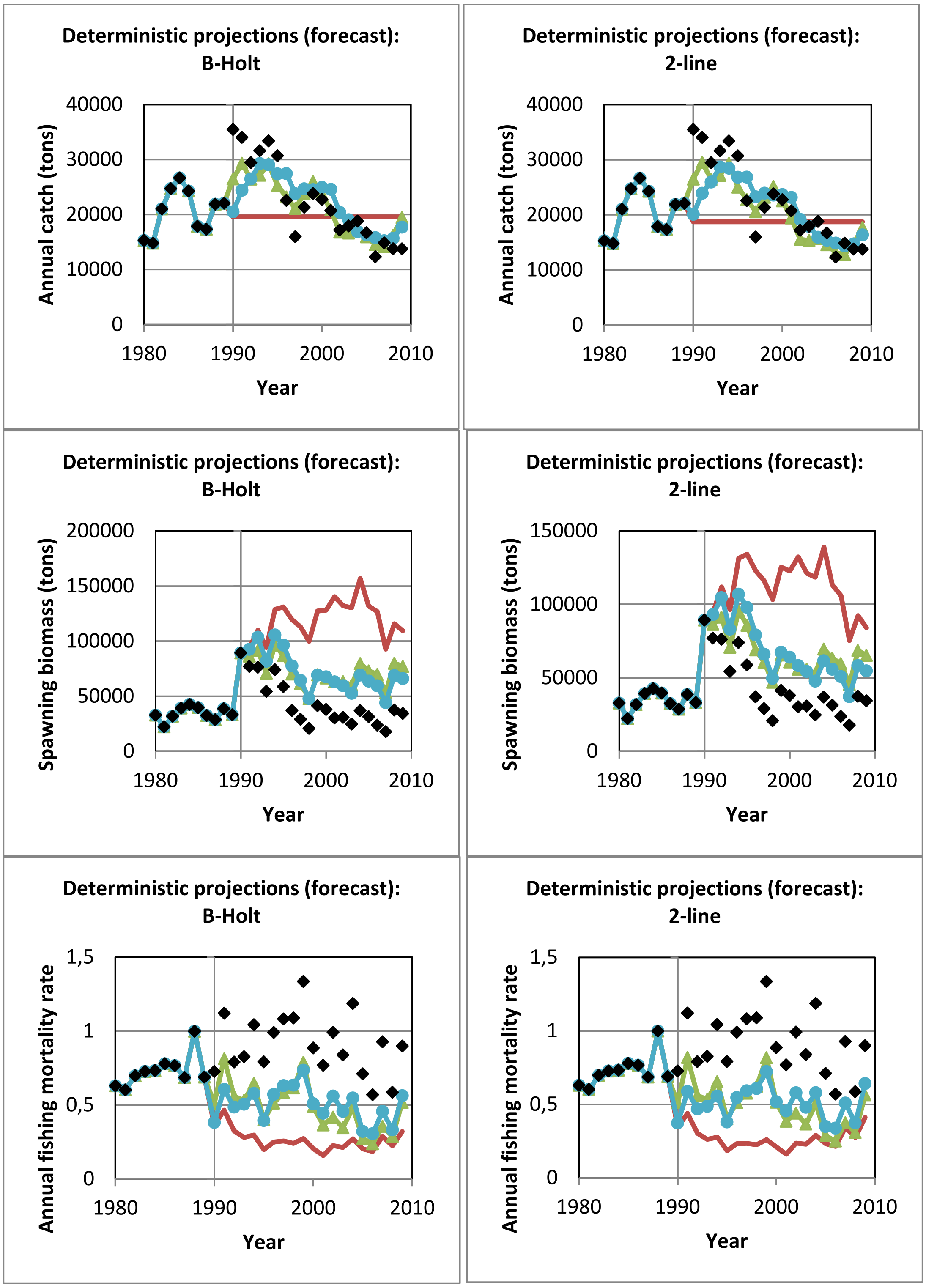


Figure 10. Deterministic "hindsight" projections with a Beverton–Holt (left) and 2-line (right) stock–recruit relationship from 1990 for a constant catch strategy (line), and for BTS-Isis survey slope (triangles) and target (dots) "forecast" MPs compared to the adjusted 2010 XSA assessment estimates for Sole in Subarea IV (black diamonds). Top two plots: total annual catch. Middle two plots: spawning biomass. Bottom two plots: annual fishing mortality.

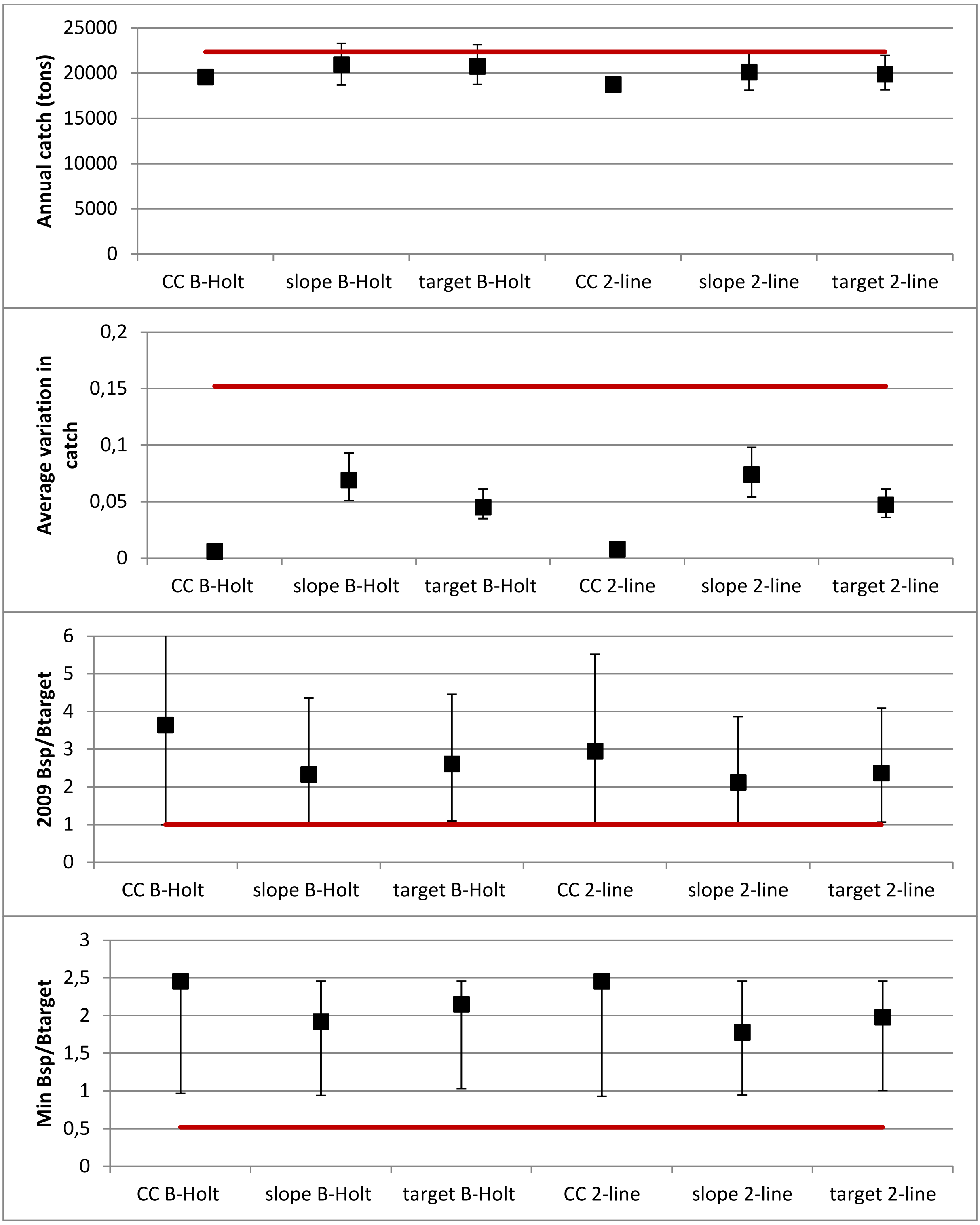


Figure 11a. Comparison of performance statistics for the "forecast" MPs for Sole in Subarea IV over the 1990 to 2009 period when projecting with a Beverton–Holt and 2-line stock–recruitment relationship: medians and 95% probability intervals of 1000 simulations. From top to bottom: average annual future catch, average interannual variation in catch, final spawning biomass as a fraction of target, and minimum future spawning biomass as a fraction of the target value. The solid horizontal lines indicate results for the adjusted 2010 XSA assessment estimates.

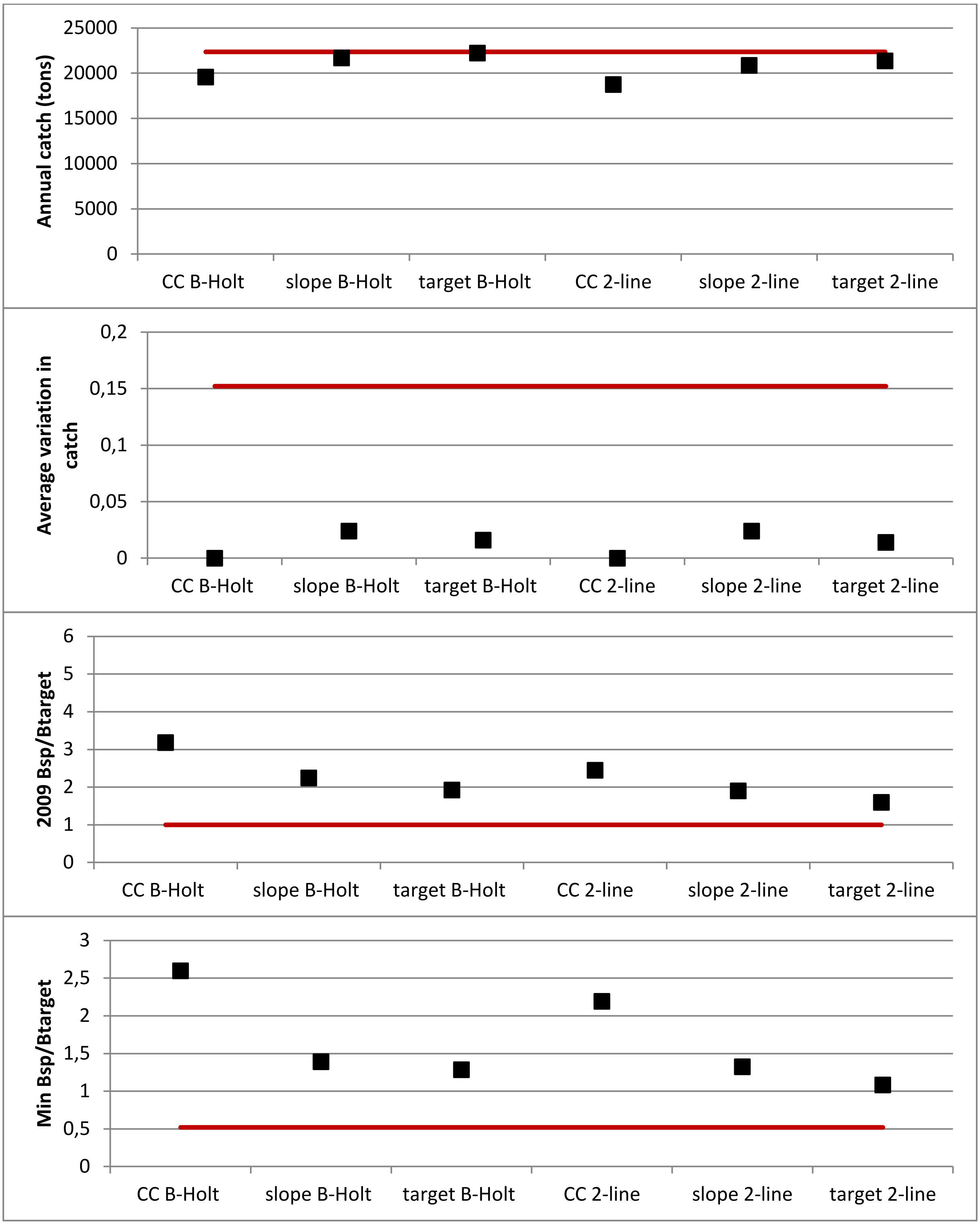


Figure 11b. Comparison of performance statistics for deterministic ("hindsight") projections under "forecast" MPs for Sole in Subarea IV over the 1990 to 2009 period with a Beverton–Holt and 2-line stock–recruitment relationship. From top to bottom: average annual future catch, average interannual variation in catch, final spawning biomass as a fraction of target, and minimum future spawning biomass as a fraction of the target value. The solid horizontal lines indicate results for the adjusted 2010 XSA assessment estimates.

APPENDIX A: (to WD7)

Input to projections

For purposes of this exercise, the 2010 ICES assessment outputs (ICES WGNSSK Report 2010) were used as the starting points from which the projections are performed.

The natural mortality rate and maturity ogive used in the 2010 XSA assessments, and as assumed here for the retrospective projections commencing in 1990, are given in Tables A1.1 and A2.1 for the North Sea Plaice and Sole stocks (subarea IV) respectively.

The annual catches for each of these stocks, the 2010 XSA estimated number of recruits (1-yr-olds) and the associated spawning biomasses are given in Tables A1.3 and A2.3 respectively. The corresponding plots of total annual catches of North Sea Plaice and Sole in Subarea IV are shown in Figures A1.1 and A2.1, while the plots for different biomass components are given in Figures A1.5 and A2.5 respectively.

Plusgroup

The "historic" population numbers and fishing mortalities-at-age for North Sea Plaice and Sole (Subarea IV) from 1957 to 1989 are taken from the XSA assessment and are assumed to be known exactly. However, difficulties arise from the manner in which the plusgroup was treated in 2010 assessments which, although this makes little difference to the overall assessment results, in mathematically inconsistent in not respecting equation (A.1) below for the dynamics. Because of the need for comparable consistent reflection of the dynamics in the alternative projections considered in this analysis, the plusgroup numbers and fishing mortalities needed to be re-estimated for the assessment in a way that avoided this inconsistency. Thus, for the sake of consistency between the XSA assessment estimates and the projections, the plusgroup numbers, $N_{y,10}$, were re-estimated such that:

$$N_{y+1,m} = \sum_{a=m-1}^{m} N_{y,a} e^{-Z_{y,a}} \tag{A.1}$$

where

$N_{y+1,m}$ is the plusgroup number of fish ($m = 10$) at the start of year *y+1*, and

$Z_{y,a} = M_{y,a} + F'_{y,a}$ is the total mortality on fish in year *y*, where

$M_{y,a}$ is the natural mortality rate, assumed to be age and year-independent, and

$F'_{y,a} = S_{y,a} F_y$ are the fishing mortalities-at-age in year *y*.

In the above equation both $N_{y,m-1}$ and $Z_{y,m-1}$ are known from the XSA assessment. In order to re-compute the plusgroup numbers for the next year, the plusgroup number for year $y-1$ needs to be known, which in turn is computed from the previous year's plusgroup number, etc. Therefore, only an estimate of the first plusgroup, $N_{1,m}$, is required to be able to compute all subsequent plusgroup numbers.

The plusgroup fishing mortality rates, $F_{y,10}$, were re-estimated using the Baranov catch equation:

$$\hat{C}_{y,m} = F'_{y,m} N_{y,m} (1 - e^{-Z_{y,m}}) / Z_{y,m} \tag{A.2}$$

In addition, flat fishing selectivity was assumed at older ages so that:

$$F'_{y,m} = F'_{y,m-1} \tag{A.3}$$

Since the above equations cannot be satisfied simultaneously, $N_{1,m}$ and $F_{y,10}$ were estimated in terms of their maximum likelihood values. The likelihood is calculated assuming that the observed plusgroup catches defined by equation (A.2) are lognormally distributed about their expected values:

$$C_{y,m} = \hat{C}_{y,m} e^{\zeta_y} \tag{A.4}$$

where $\zeta_y \sim N(0, (\sigma^C)^2)$. Similarly, the plusgroup fishing mortalities are assumed to be lognormally distributed about their expected values:

$$F'_{y,m} = F'_{y,m-1} e^{\tau_y} \tag{A.5}$$

where $\tau_y \sim N(0, (\sigma^F)^2)$.

The contributions to the negative of the (penalised) log-likelihood function are given by:

$$-\ln L = -\ln L^F - \ln L^C \tag{A.6}$$

where

$$-\ln L^F = \sum_y [\ln \sigma^F + (\ln F'_{y,m-1} - \ln F'_{y,m})^2 / 2(\sigma^F)^2] \tag{A.7}$$

and

$$-\ln L^C = \sum_y [\ln \sigma^C + (\ln C_{y,m} - \ln \hat{C}_{y,m})^2 / 2(\sigma^C)^2] \tag{A.8}$$

where σ^F and σ^C are the standard deviation of the residuals, estimated in the fitting procedure by their maximum likelihood values

$$\sigma^F = \sqrt{1/n \sum_y (\ln F'_{y,m} - \ln F'_{y,m-1})^2} \quad \text{(A.9)}$$

and

$$\sigma^C = \sqrt{1/n \sum_y (\ln C_{y,m} - \ln \hat{C}_{y,m})^2} \quad \text{(A.10)}$$

where n is the number of years over which the summation is taken.

The adjusted population numbers and fishing mortality matrices, $N_{y,a}$ and $F'_{y,a}$, are given in Tables A1.4 and A1.5 for North Sea Plaice, and Tables A2.4 and A2.5 for Sole in Subarea IV. Due to the near-zero estimates of σ^F, the $F'_{y,a}$ matrices remain effectively unchanged from those estimated in the 2010 ICES assessments. The plots of the adjusted plusgroup population numbers, $N_{y,10}$, and annual plusgroup catches (landings and discards), $C_{y,10}$, are shown in Figures A1.2 and A1.3 for North Sea Plaice, and Figures A2.2 and A2.3 for Sole in subarea IV.

The catch and population weights-at-age matrices, $w^C_{y,a}$ and $w^S_{y,a}$ were taken directly from those used in the 2010 XSA assessments, shown in Figures A1.13 and A1.14 for North Sea Plaice and A2.11 and A2.12 for Sole in Subarea IV. Decreasing trends in weights at older ages are clearly visible in these plots since 1990 for both Plaice and Sole stocks.

The age- and year-dependent fishing selectivities were derived from the adjusted $F'_{y,a}$ matrix such that

$$S_{y,a} = F'_{y,a} / F_y \quad \text{(A.11)}$$

where $F_y = \max_a (F'_{y,a})$.

The annual fishing selectivity-at-age vectors are shown in Figures A1.7 to A1.10 for North Sea Plaice, and Figures A2.7 and A2.8 for Sole caught in subarea IV.

Stock–recruitment relationship

The number of recruits is assumed to be lognormally distributed about a stock–recruitment relationship such that

$$R_y^{XSA} = R_y e^{\varsigma_y} \quad \text{(A.12)}$$

where

R_y^{XSA} are the number of recruits in year y, input from the 2010 XSA assessment,

R_y is the number of recruits according to some stock–recruit relationship, and

ς_y are the corresponding recruitment residuals.

The objective function minimized to estimate the parameters of the relationship is given by

$$-\ln L = \sum_y [\ln \sigma^R + (\ln R_y^{XSA} - \ln R_y)^2 / 2(\sigma^R)^2] \quad \text{(A.13)}$$

where $\sigma^R = \sqrt{1/n\sum_y(\ln R_y^{XSA} - \ln R_y)^2}$ is the standard deviation of the residuals estimated in the fitting procedure by its maximum likelihood value and y runs over the "historic" years from 1957 to 1989 for the stochastic ("forecast") projections, and from 1957 to 2009 for the deterministic ("hindsight") projections.

Two forms of relationships are considered.

Beverton–Holt:

The number of recruits is given by a Beverton–Holt stock–recruitment relationship such that

$$R_y = \frac{\alpha\, B_{y-1}^{sp}}{\beta + B_{y-1}^{sp}} \qquad \text{(A14)}$$

where

> B_{y-1}^{sp} is the spawning biomass in year $y-1$, corresponding to the adjusted 2010 XSA assessment estimates, and
>
> α and β are the stock–recruitment parameters which are estimated.

Note: The "steepness" of the stock–recruitment curves (recruitment at $B^{sp} = 0.2K$ as a fraction of recruitment at $B^{sp} = K$) was estimated to be close to one, i.e. $\beta = 0$ and hence effectively constant recruitment regardless of the level of spawning biomass. This is frequently criticized as there is negligible penalty if harvests reduce the resource to very low levels. Therefore, the steepness parameter, h, was fixed to 0.9 when estimating α , with β given in terms of h such that

$$\beta = \alpha(SBR)(1-h)/4h \qquad \text{(A.15)}$$

where SBR is the pre-exploitation spawning biomass per recruit.

Plots of the number of recruits obtained from the XSA assessment, along with the corresponding Beverton–Holt estimates, are shown in Figures A1.6 and A1.7 for Plaice and A2.6 and A2.7 for Sole respectively.

2-Line:

Due to the unrealistically high estimates of h for the Beverton–Holt relationship ($h = 0.98$ for North Sea Plaice and $h = 1$ for Sole in Subarea IV), an alternative stock–recruit relation was tested: a two line (or "hockey-stick") stock–recruit function, where the expected number of recruits is constant above a certain spawning biomass level, and as the spawning biomass falls below that level, the number of recruits decreases linearly to zero. The level chosen is the minimum spawning biomass (B0) in the time-series from the (adjusted) XSA assessment.

The number of recruits, R_y, is given by

$$\begin{aligned} B^{sp}_{y-1} < B^0: \quad & R_y = (\alpha B^{sp}_{y-1} / B^0) e^{\zeta_y} \\ B^{sp}_{y-1} \geq B^0: \quad & R_y = \alpha e^{\zeta_y} \end{aligned} \tag{A.16}$$

where

α is the number of recruits (constant) when the spawning biomass is above a prespecified minimum value,

$B^0 = \min_y (B^{XSA}_y)$ denotes the minimum spawning biomass over the period under consideration below which the number of recruits decline linearly.

The stock–recruit parameter estimates obtained by minimizing equation (A.13) are given in Table A1.2 for North Sea Plaice and A2.2 for Sole in Subarea IV.

Survey abundance data

A variety of age-disaggregated survey data were used to tune the XSA assessment. However, for the purposes of this paper an age-aggregated index is required:

$$I^i_y = \sum_a w^S_{y,a} I^i_{y,a} \tag{A.17}$$

where $I^i_{y,a}$ are the age-disaggregated survey indices corresponding to BTS-Isis, BTS-Tridens and SNS for Plaice, and BTS-Isis and SNS for Sole. The abundance indices, I^i_y, are assumed to be lognormally distributed about their expected values such that

$$I^i_y = \hat{I}_y e^{\varepsilon^i_y} \tag{A.18}$$

where

I^i_y is the age-aggregated survey abundance index *i* for year *y* given by equation (A.17),

$\hat{I}^i_y = q^i B^{sur\_i}_y$ is the corresponding model estimate, where

$B^{sur\_i}_y$ is the survey biomass estimate for year y,

q^i is the constant of proportionality for abundance series *i* given by

$$\ln q^i = 1/n \sum_y (\ln I^i_y - \ln B^{sur\_i}_y) \tag{A.19}$$

and ε^i_y are the residuals

$$\varepsilon^i_y = \ln I^i_y - \ln(q^i B^{sur\_i}_y) \tag{A.20}$$

with the standard deviation of the residuals for survey index i given by

$$\sigma^i = \sqrt{1/n(\sum_y \varepsilon_y^i)^2} \qquad \text{(A.21)}$$

The biomass for year y corresponding to survey index i is given by

$$B_y^{sur\_i} = \sum_{a=1}^{10} S_a^{sur\_i} w_{y,a}^S N_{y,a} \qquad \text{(A.22)}$$

where

$w_{y,a}^S$ denote the population weights-at-age for each year which are input,

$N_{y,a}$ are the adjusted 2010 XSA assessment population numbers-at-age, and

$S_a^{sur\_i}$ is the fishing selectivity vector associated with survey abundance index i given by

$$S_a^{sur\_i} = 1/n \sum_y I_{y,a}^{sur\_i} / N_{y,a} \qquad \text{(A.23)}$$

where

$I_{y,a}^{sur\_i}$ is the age-disaggregated survey data matrix corresponding to index i which is input,

n is the number of years of survey data in index i, and

$N_{y,a}$ corresponds to the 2010 XSA population numbers with adjusted plus-group.

The age-aggregated survey estimates are given in Tables A1.6 and A2.6 for North Sea Plaice and Sole respectively. The corresponding plots of the survey indices are shown in Figures A1.4 and A2.4.

Section A1: North Sea Plaice (subarea IV)

The natural mortality-at-age and maturity-at-age vectors used in the XSA assessment and retrospective projections from 1990 for North Sea Plaice (subarea IV).

Table A1.1. Natural mortality-at-age and maturity-at-age vectors.

AGE	1	2	3	4	5	6	7	8	9	10
Natural mortality rate	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
Maturity	0	0.5	0.5	1.0	1.0	1.0	1.0	1.0	1.0	1.0

The stock–recruit parameters estimated for the Beverton–Holt and 2-line functions used in the deterministic and stochastic projections.

Table A1.2. Stock–recruit parameters

		BEVERTON–HOLT		2_LINE	
	Period	α (thousands)	β (tons)	α (thousands)	B^0 (tons)
Determinstic ("Hindsight")	1957–2009	2301950	434656	927286	198132
Stochastic ("Forecast")	1957–1989	2147080	405408	939615	250267

Year	Number of recruits	Spawning bio-mass	Total catch	Landings	Discards
1957	457973	288705	78443	70563	7880
1958	698110	291614	88191	73354	14837
1959	863386	291514	109164	79300	29864
1960	757298	302878	117334	87541	29793
1961	860576	309561	118474	85984	32490
1962	589154	365078	125375	87472	37903
1963	688366	348818	148376	107118	41258
1964	2231500	339871	147571	110540	37031
1965	694573	316015	140223	97143	43080
1966	586777	337571	166552	101834	64718
1967	401295	403512	163365	108819	54546
1968	434277	382406	139521	111534	27987
1969	648869	350206	142820	121651	21169
1970	650576	326437	159982	130342	29640
1971	410270	291251	136939	113944	22995
1972	366617	299224	142475	122843	19632
1973	1312009	252552	143783	130429	13354
1974	1132726	259124	157485	112540	44945
1975	864773	273132	195235	108536	86699
1976	692682	292629	166917	113670	53247
1977	988665	307704	176689	119188	57501
1978	912345	295538	159639	113984	45655
1979	891239	287895	213282	145347	67935
1980	1128156	263884	171031	139951	31080
1981	865944	252209	172778	139747	33031
1982	2031170	250267	203674	154547	49127
1983	1308491	303768	218521	144038	74483
1984	1259358	314454	226963	156147	70816
1985	1848419	337665	220387	159838	60549
1986	4760609	364215	295300	165347	129953
1987	1962845	442388	344194	153670	190524
1988	1770461	382424	310898	154475	156423
1989	1186811	411792	277611	169818	107793
1990	1036516	371947	227465	156240	71225
1991	914585	343770	228939	148004	80935
1992	776744	279797	182239	125190	57049
1993	530684	242006	152129	117113	35016
1994	442947	209421	134177	110392	23785
1995	1164164	201208	120184	98356	21828
1996	1290364	202807	133722	81673	52049
1997	2155842	211554	183193	83048	100145
1998	774928	228808	175285	71534	103751
1999	840878	201461	151638	80662	70976
2000	991191	228618	125459	81148	44311
2001	540350	262660	182272	81963	100309
2002	1726207	198132	124607	70217	54390
2003	537804	230789	144294	66502	77792
2004	1248173	215963	115902	61436	54466
2005	791655	252773	109576	55700	53876
2006	922375	275293	119789	57943	61846
2007	1046417	271502	89179	49744	39435
2008	821795	347508	94749	48874	45875
2009	1017863	403767	100198	54973	45225

Table A1.3. Spawing biomass estimates from the adjusted 2010 XSA assessment, with total annual catches (landings and discards) for North Sea Plaice.

YEAR	1	2	3	4	5	6	7	8	9	10
1957	457973	256778	322069	182986	117504	49780	48438	35192	20763	58933
1958	698110	383614	184865	225749	122171	75186	36568	33338	23255	53959
1959	863386	568706	270362	123650	142799	76063	49331	25309	22555	50581
1960	757298	670799	377298	171551	76786	85609	46907	31440	16805	45847
1961	860576	614899	441591	239779	105744	48183	50972	28949	19875	38652
1962	589154	706789	416674	283132	151855	63044	31337	32158	16921	36179
1963	688366	484324	465009	259569	172009	89026	37245	19737	20503	32369
1964	2231500	536380	304564	276885	152215	101919	50127	21480	11359	30565
1965	694573	1956330	325547	176043	156783	80258	56631	30309	13162	23972
1966	586777	586899	1355540	198052	105458	99441	43686	33776	19288	22299
1967	401295	494319	371937	832385	116531	59210	63824	23833	20304	24356
1968	434277	343893	314556	224454	500704	65484	32351	42364	13952	26340
1969	648869	322587	233484	201830	141578	314124	42894	19435	28723	25665
1970	650576	506081	213512	152352	129908	93520	185267	28910	11797	34472
1971	410270	471051	296427	118122	83215	74030	51104	92598	20156	26245
1972	366617	305254	305838	182003	72494	50103	45122	30153	55506	27947
1973	1312010	263017	188694	185322	108922	43137	29096	27149	16912	48925
1974	1132730	1060050	160417	110708	97545	57136	25825	17876	15198	37047
1975	864773	821976	643812	88838	59831	48609	32888	15753	10162	29077
1976	692682	548525	450535	342684	46074	29718	23712	18465	8620	20423
1977	988665	449171	330275	266210	201243	28417	17430	12780	10628	16840
1978	912345	647406	253598	182219	146168	93607	16894	10147	6787	14865
1979	891239	608629	381577	144234	102938	83416	50378	9636	5993	12245
1980	1128160	526305	290915	177429	66449	47031	37199	22538	4761	8377
1981	865944	804536	297898	135126	86149	36186	25369	20569	12348	6997
1982	2031170	655698	448153	151458	67118	43539	20882	13857	10914	10242
1983	1308490	1443460	353260	202293	69838	33268	23392	11676	7335	10880
1984	1259360	934165	777188	181001	86673	34500	17757	13351	6576	9367
1985	1848420	843888	486900	392310	89506	41365	18156	9917	6807	8150
1986	4760610	1286790	475587	269456	176694	49864	22568	9893	5502	7895
1987	1962840	3243130	633464	228453	129409	78140	22104	11575	4680	5789
1988	1770460	1432170	1546360	290743	99541	54212	37434	9107	6344	4891
1989	1186810	1270380	703021	723770	134181	44160	25017	17916	3122	4841
1990	1036520	869783	642864	351602	353191	64212	21540	12585	9019	2062
1991	914585	798177	490389	328242	185828	162794	32481	11758	7377	5991
1992	776744	651967	394198	229534	147764	93483	72635	15207	5661	6678
1993	530684	567595	339205	185748	106060	60448	51973	37617	7130	4530
1994	442947	385219	315695	167606	87929	45514	26903	34212	23315	4880
1995	1164160	340377	214579	155030	74346	42117	21652	9768	24869	19340
1996	1290360	932940	194551	101746	64911	31921	20918	10296	4685	36050
1997	2155840	1060700	488817	88535	44228	27742	15018	8839	5009	15152
1998	774928	1827460	432991	175218	38011	18319	12140	7538	4543	9902
1999	840878	601558	1009900	143943	54210	18214	10426	6242	4436	7747
2000	991191	639225	337810	544968	40139	25850	9929	6787	3512	7050
2001	540350	795939	400484	219442	274990	20887	14274	6708	5002	6871
2002	1726210	455774	350797	134668	81575	111630	12252	8670	4997	9302
2003	537804	1266050	228653	189466	64328	37549	64349	6567	6143	10916
2004	1248170	421550	612659	111405	106730	28747	18390	36518	4412	13717
2005	791655	907301	200817	346915	61062	76051	15049	12051	28709	14799
2006	922375	624505	496183	115723	217144	35145	54069	8405	8417	36159
2007	1046420	624515	334116	282853	72005	153819	25328	43503	5468	34097
2008	821795	872142	352979	203537	202618	50212	116243	20417	36033	32071
2009	1017860	614491	539583	245324	143705	153734	39699	90345	16765	60403

Table A1.4. Population numbers-at-age for North Sea Plaice taken from 2010 ICES XSA assessment, but with adjusted plusgroup as discussed in text.

Year	1	2	3	4	5	6	7	8	9	10
1957	0.077	0.229	0.255	0.304	0.347	0.208	0.274	0.314	0.290	0.290
1958	0.105	0.250	0.302	0.358	0.374	0.321	0.268	0.291	0.323	0.323
1959	0.152	0.310	0.355	0.376	0.412	0.383	0.350	0.309	0.367	0.367
1960	0.108	0.318	0.353	0.384	0.366	0.419	0.383	0.359	0.383	0.383
1961	0.097	0.289	0.344	0.357	0.417	0.330	0.361	0.437	0.381	0.381
1962	0.096	0.319	0.373	0.398	0.434	0.426	0.362	0.350	0.395	0.395
1963	0.149	0.364	0.418	0.434	0.423	0.474	0.450	0.452	0.448	0.448
1964	0.032	0.399	0.448	0.469	0.540	0.488	0.403	0.390	0.459	0.459
1965	0.068	0.267	0.397	0.412	0.355	0.508	0.417	0.352	0.410	0.410
1966	0.071	0.356	0.388	0.430	0.477	0.343	0.506	0.409	0.435	0.435
1967	0.054	0.352	0.405	0.408	0.476	0.504	0.310	0.435	0.428	0.428
1968	0.197	0.287	0.344	0.361	0.366	0.323	0.410	0.289	0.351	0.351
1969	0.149	0.313	0.327	0.341	0.315	0.428	0.295	0.399	0.356	0.356
1970	0.223	0.435	0.492	0.505	0.462	0.504	0.594	0.261	0.467	0.467
1971	0.196	0.332	0.388	0.388	0.407	0.395	0.428	0.412	0.407	0.407
1972	0.232	0.381	0.401	0.413	0.419	0.443	0.408	0.478	0.434	0.434
1973	0.113	0.394	0.433	0.542	0.545	0.413	0.387	0.480	0.475	0.475
1974	0.221	0.399	0.491	0.515	0.596	0.452	0.394	0.465	0.486	0.486
1975	0.355	0.501	0.531	0.557	0.600	0.618	0.477	0.503	0.553	0.553
1976	0.333	0.407	0.426	0.432	0.383	0.434	0.518	0.452	0.445	0.445
1977	0.323	0.472	0.495	0.500	0.665	0.420	0.441	0.533	0.514	0.514
1978	0.305	0.429	0.464	0.471	0.461	0.520	0.461	0.427	0.470	0.470
1979	0.427	0.638	0.666	0.675	0.683	0.708	0.704	0.605	0.678	0.678
1980	0.238	0.469	0.667	0.622	0.508	0.517	0.492	0.502	0.530	0.530
1981	0.178	0.485	0.576	0.600	0.582	0.450	0.505	0.534	0.536	0.536
1982	0.242	0.518	0.695	0.674	0.602	0.521	0.481	0.536	0.565	0.565
1983	0.237	0.519	0.569	0.748	0.605	0.528	0.461	0.474	0.565	0.565
1984	0.300	0.552	0.584	0.604	0.640	0.542	0.482	0.574	0.571	0.571
1985	0.262	0.473	0.492	0.698	0.485	0.506	0.507	0.489	0.539	0.539
1986	0.284	0.609	0.633	0.633	0.716	0.714	0.568	0.648	0.739	0.739
1987	0.215	0.641	0.679	0.731	0.770	0.636	0.787	0.501	0.661	0.661
1988	0.232	0.612	0.659	0.673	0.713	0.673	0.637	0.971	0.742	0.742
1989	0.211	0.581	0.593	0.617	0.637	0.618	0.587	0.586	1.251	1.251
1990	0.161	0.473	0.572	0.538	0.675	0.582	0.505	0.434	0.515	0.515
1991	0.238	0.605	0.659	0.698	0.587	0.707	0.659	0.631	0.594	0.594
1992	0.214	0.553	0.652	0.672	0.794	0.487	0.558	0.657	0.902	0.902
1993	0.220	0.487	0.605	0.648	0.746	0.710	0.318	0.378	0.771	0.771
1994	0.163	0.485	0.611	0.713	0.636	0.643	0.913	0.219	0.277	0.277
1995	0.121	0.459	0.646	0.771	0.745	0.600	0.643	0.635	0.104	0.104
1996	0.096	0.546	0.687	0.733	0.750	0.654	0.761	0.621	0.889	0.889
1997	0.065	0.796	0.926	0.746	0.781	0.726	0.589	0.566	0.611	0.611
1998	0.153	0.493	1.001	1.073	0.636	0.464	0.565	0.430	0.523	0.523
1999	0.174	0.477	0.517	1.177	0.641	0.507	0.329	0.475	0.447	0.447
2000	0.119	0.368	0.331	0.584	0.553	0.494	0.292	0.205	0.330	0.330
2001	0.070	0.719	0.990	0.890	0.802	0.433	0.399	0.195	0.144	0.144
2002	0.210	0.590	0.516	0.639	0.676	0.451	0.524	0.245	0.170	0.170
2003	0.144	0.626	0.619	0.474	0.705	0.614	0.467	0.298	0.118	0.118
2004	0.219	0.642	0.469	0.501	0.239	0.547	0.323	0.141	0.103	0.103
2005	0.137	0.504	0.451	0.369	0.452	0.241	0.482	0.259	0.085	0.085
2006	0.290	0.525	0.462	0.374	0.245	0.228	0.117	0.330	0.168	0.168
2007	0.082	0.471	0.396	0.234	0.260	0.180	0.116	0.088	0.110	0.110
2008	0.191	0.380	0.264	0.248	0.176	0.135	0.152	0.097	0.020	0.020
2009	0.168	0.426	0.257	0.204	0.184	0.129	0.086	0.087	0.035	0.035

Table A1.5. Fishing mortality-at-age for North Sea Plaice taken from the 2010 ICES XSA assessment with adjusted plusgroup.

Year	BTS-Isis	BTS-Tridens	SNS
1970			2272.52
1971			4115.48
1972			3419.04
1973			3490.6
1974			2731.37
1975			3671.22
1976			1302.31
1977			2276.69
1978			2921.2
1979			3498.43
1980			5141.19
1981			3844.22
1982			4781.09
1983			3369.44
1984			4034.47
1985	45.26		3741.39
1986	62.71		5260.06
1987	98.30		4911.93
1988	74.58		4979.08
1989	72.71		3975.19
1990	47.12		2442.32
1991	48.61		4499.28
1992	47.21		4138.53
1993	60.21		2466.78
1994	33.59		1901.9
1995	26.54		1732.16
1996	46.13	5.09	2485.38
1997	48.85	6.67	3986.73
1998	58.91	9.23	4766.81
1999	51.74	11.05	4452.05
2000	31.36	10.80	1576.88
2001	30.28	8.65	1130.47
2002	36.30	10.54	1665.43
2003	30.31	15.94	
2004	33.38	16.20	1229.38
2005	21.25	16.83	759.74
2006	18.63	20.29	909.908
2007	33.02	21.85	897.587
2008	36.64	37.93	1104.63
2009	51.01	37.52	1098.33

Table A1.6. Age-aggregated survey biomass indices for North Sea Plaice.

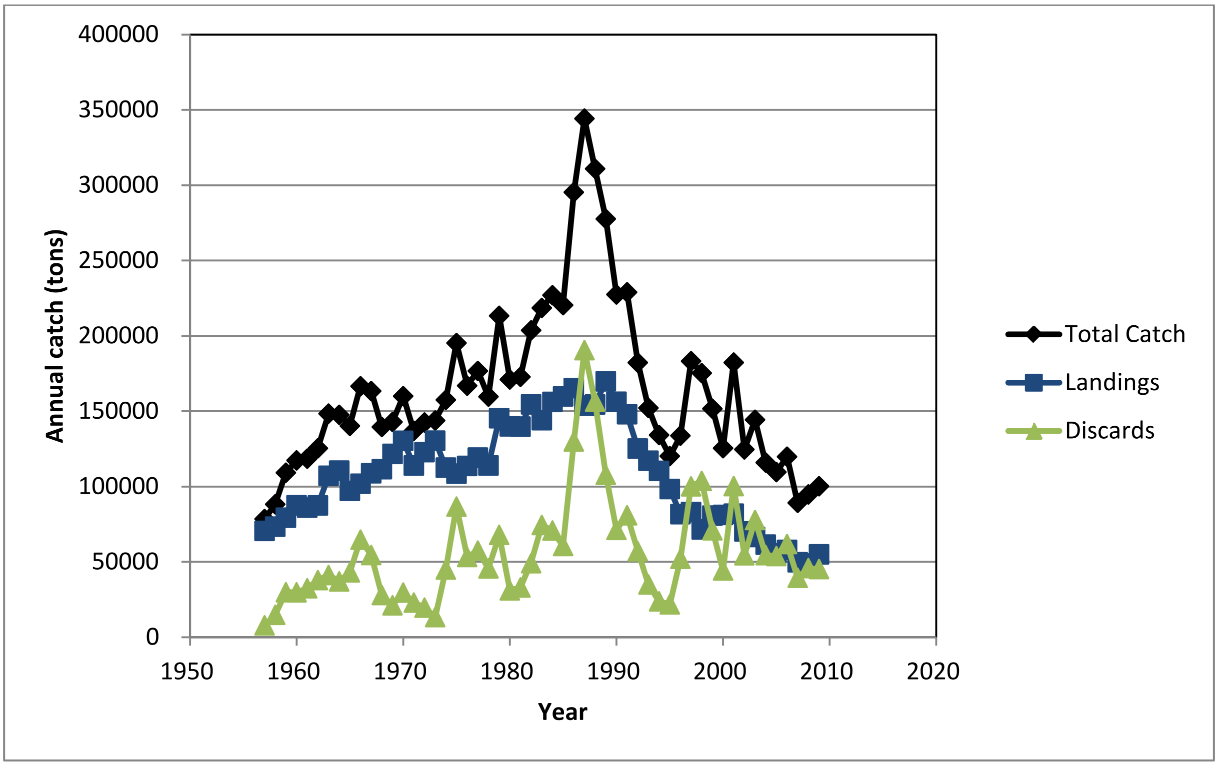


Figure A1.1. Total annual catch of North Sea Plaice in tons consisting of landings plus discards.

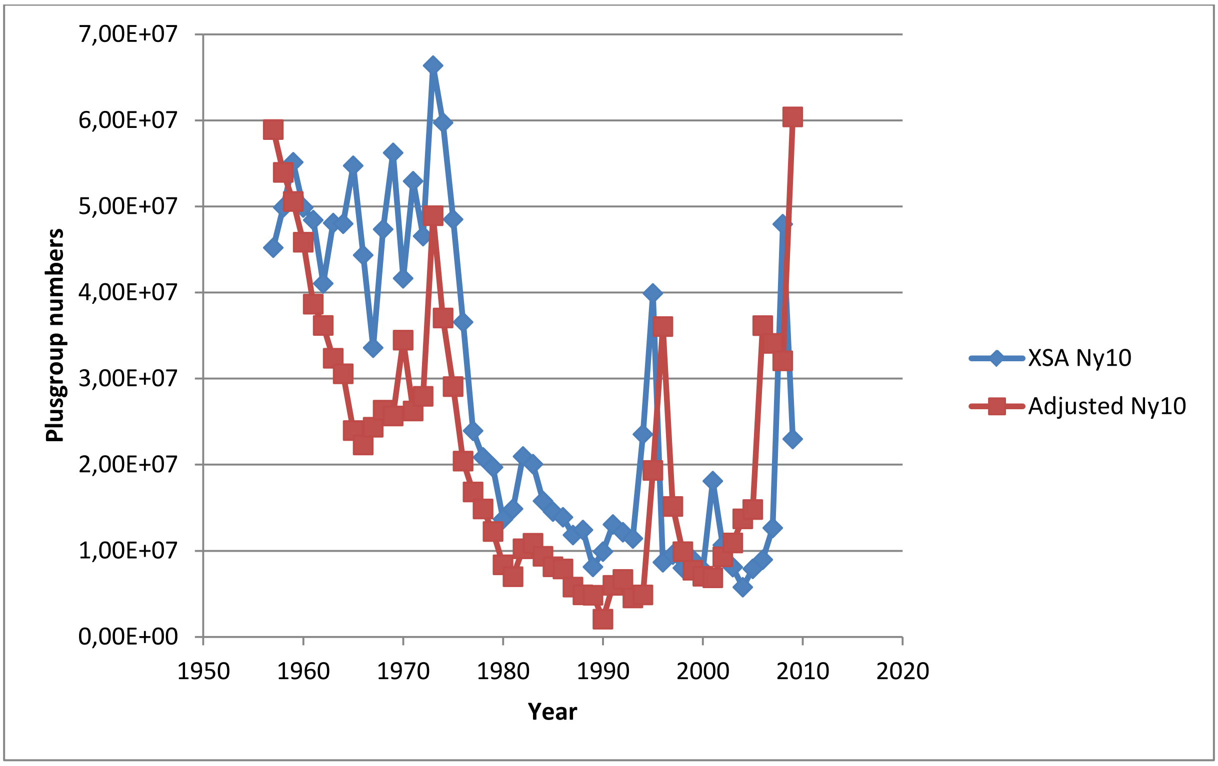


Figure A1.2. Adjusted plusgroup population numbers compared to the XSA estimates for North Sea Plaice.

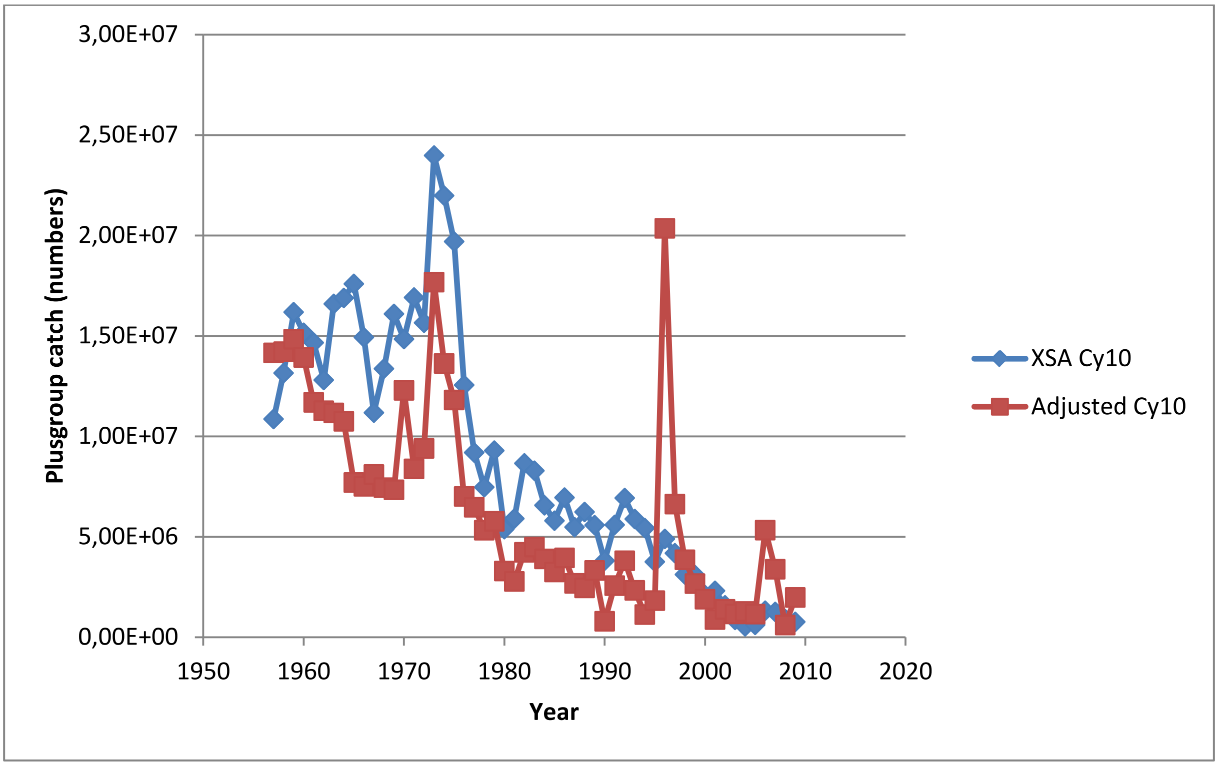


Figure A1.3. Adjusted plusgroup catch compared to observed catch for North Sea Plaice.

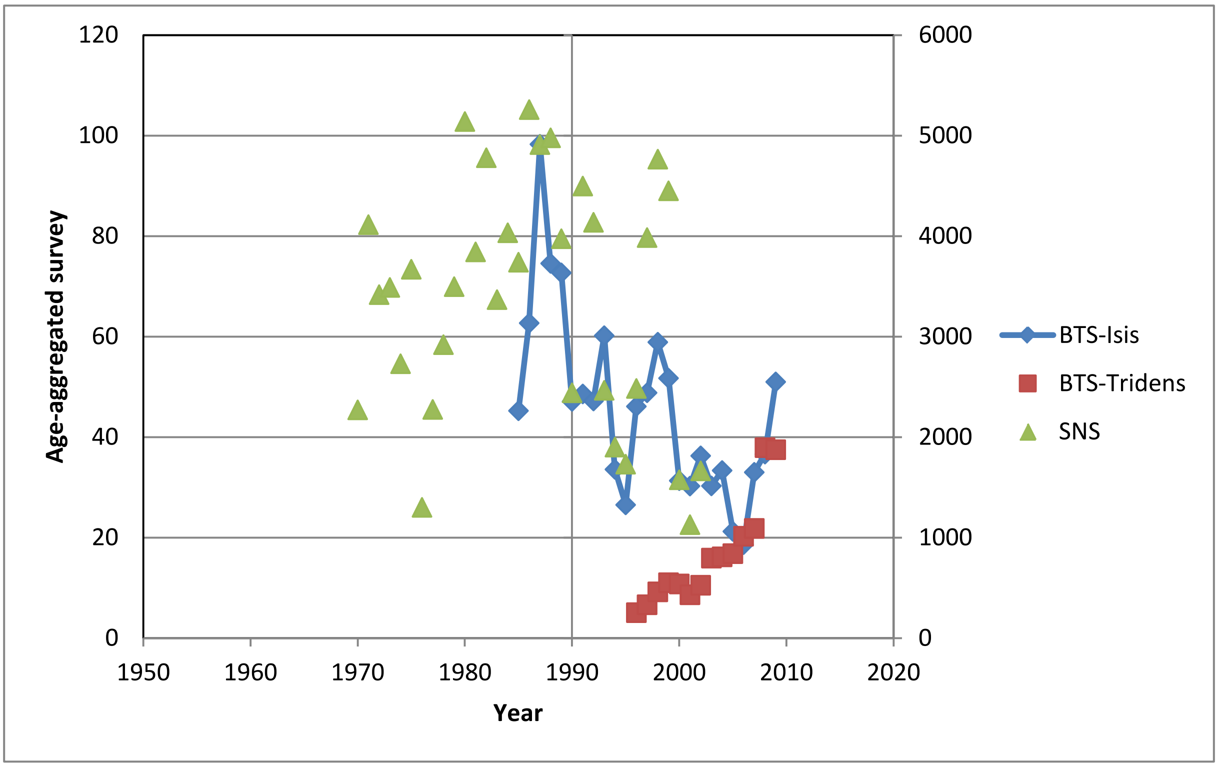


Figure A1.4. Age-aggregated survey series for North Sea Plaice.

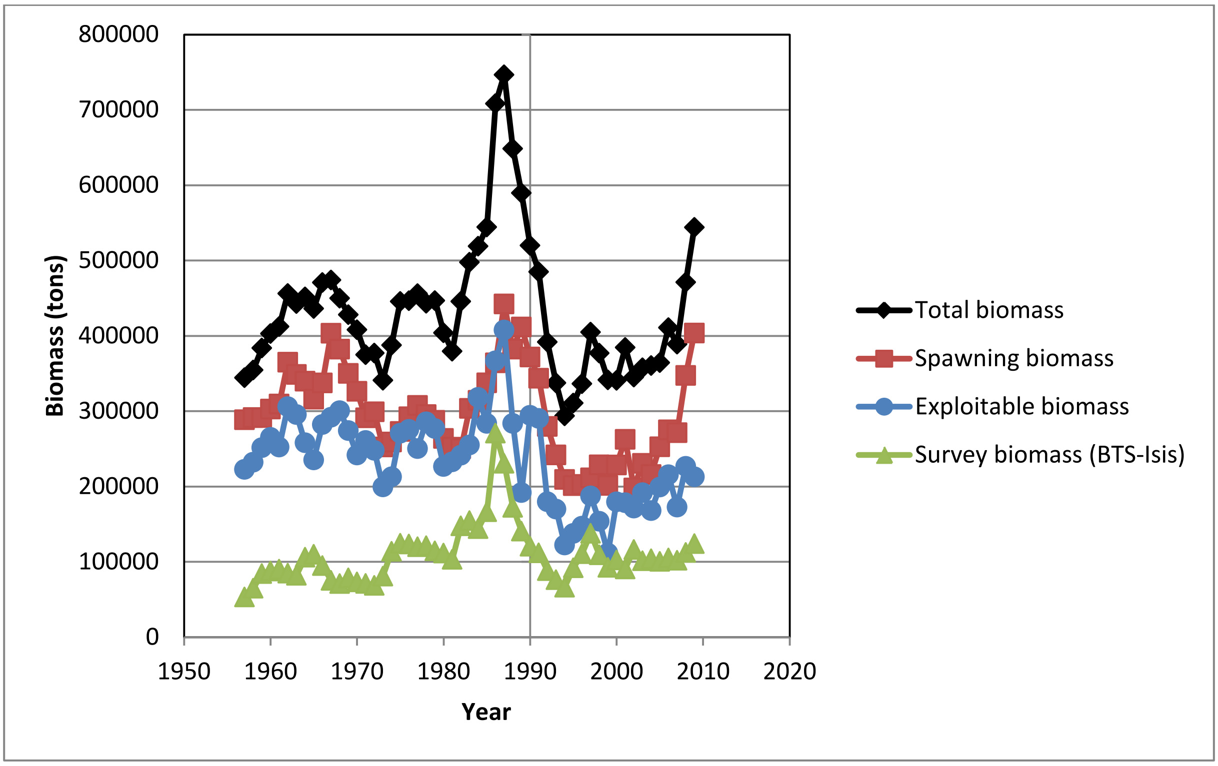


Figure A1.5. Trajectories for various biomass components from the 2010 XSA assessment with the plusgroup adjusted as detailed in text for North Sea Plaice.

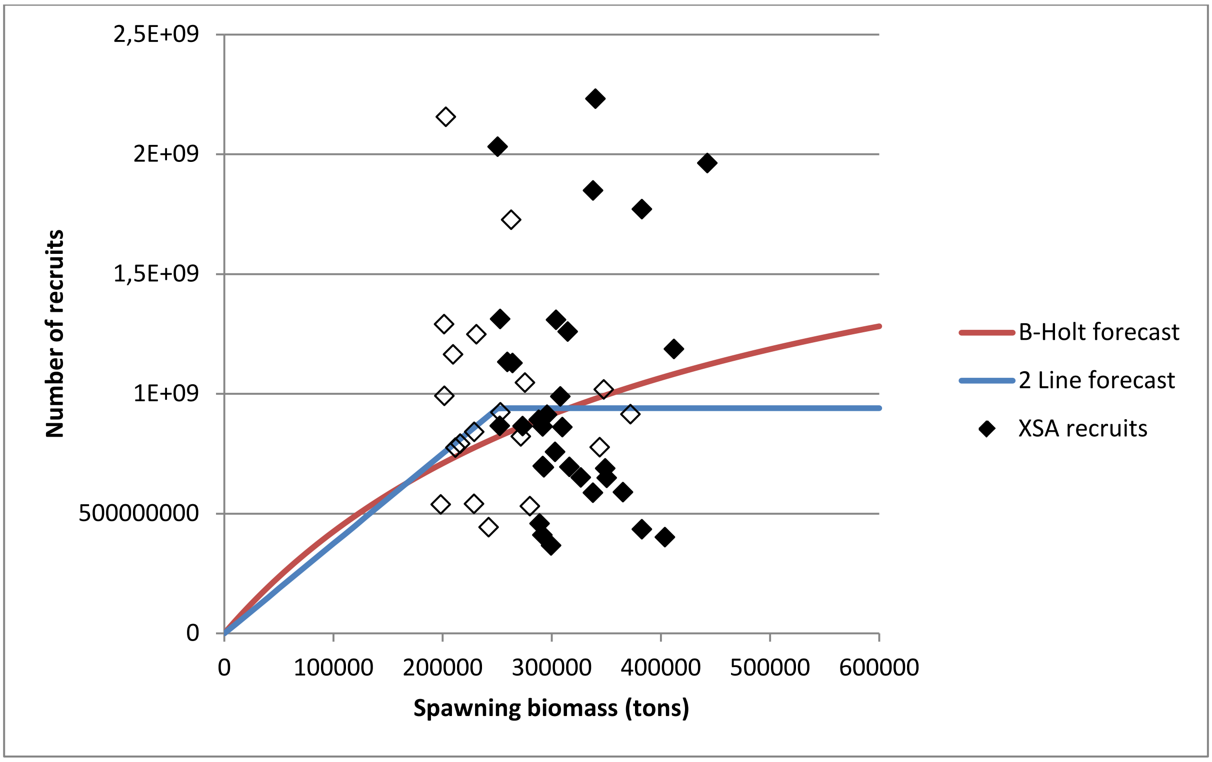


Figure A1.6. Number of recruits (1-yr-olds) estimated in the 2010 XSA assessment for North Sea Plaice (diamonds) compared to the number of recruits in terms of a Beverton–Holt stock-recruitment curve when fixing *h* to 0.9 and a 2-line stock recruit relationship fitted to data from 1957 to 1989 (forecast). Recruitments from 1990 onwards are shown by open diamonds.

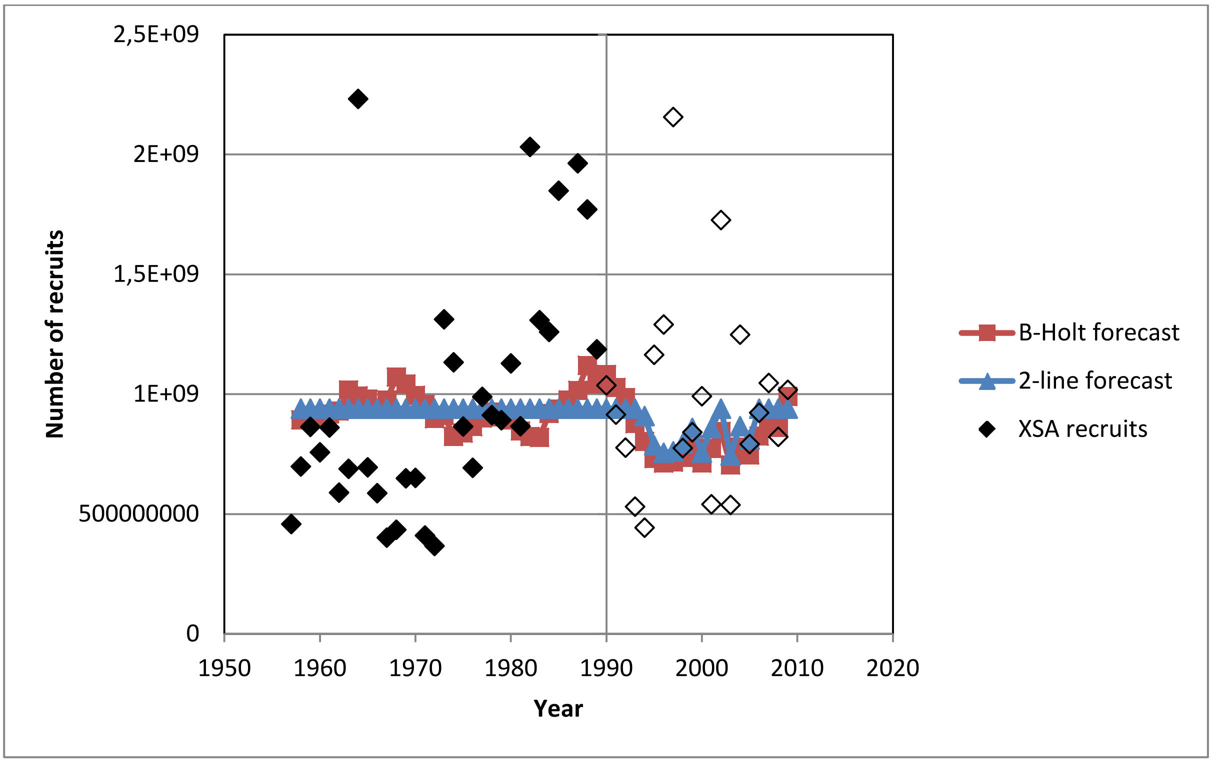


Figure A1.7. Annual number of recruits (1-yr-olds) estimated in the 2010 XSA assessment for North Sea Plaice (diamonds) compared to the annual number of recruits in terms of a Beverton–Holt stock–recruitment curve fixing *h*=0.9 (squares) and a 2-line stock–recruit relationship (triangles).

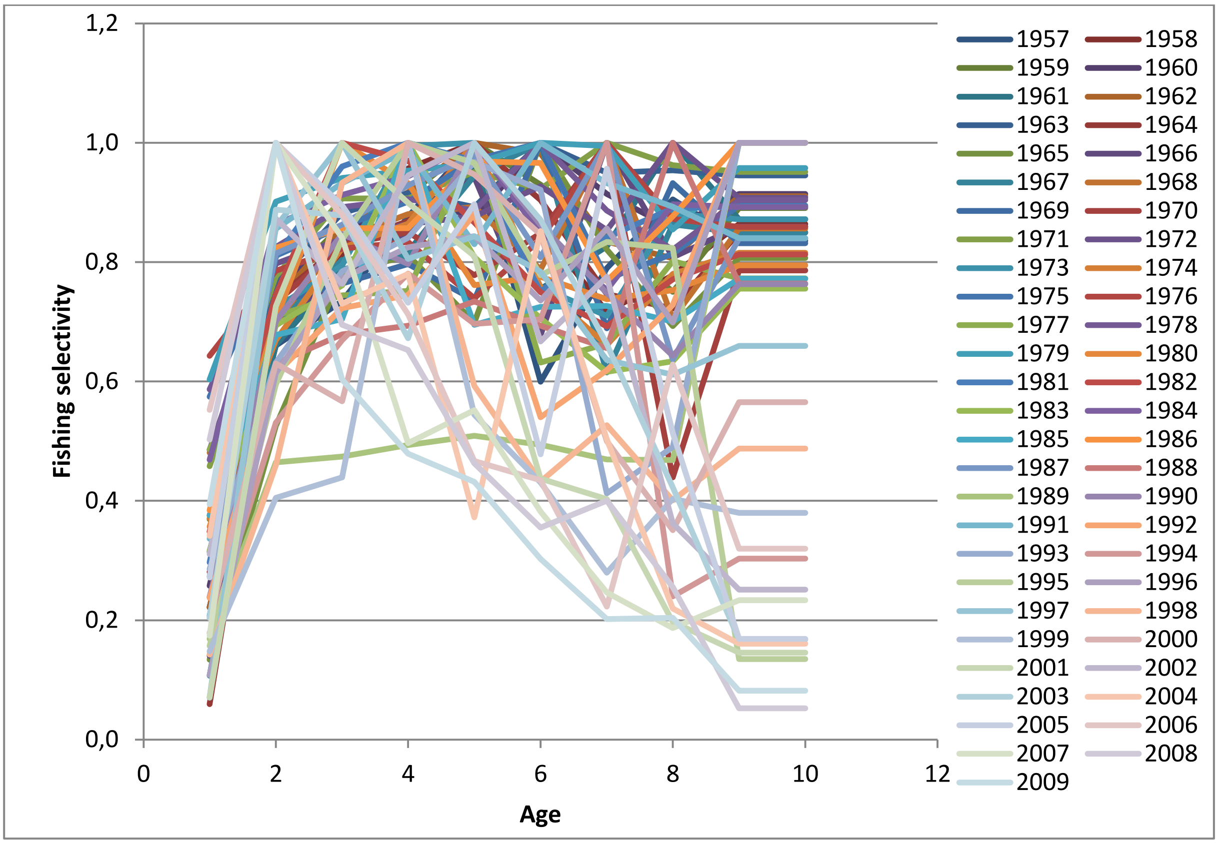


Figure A1.8. Fishing selectivities-at-age over the assessment period from 1957 to 2009 for North Sea Plaice.

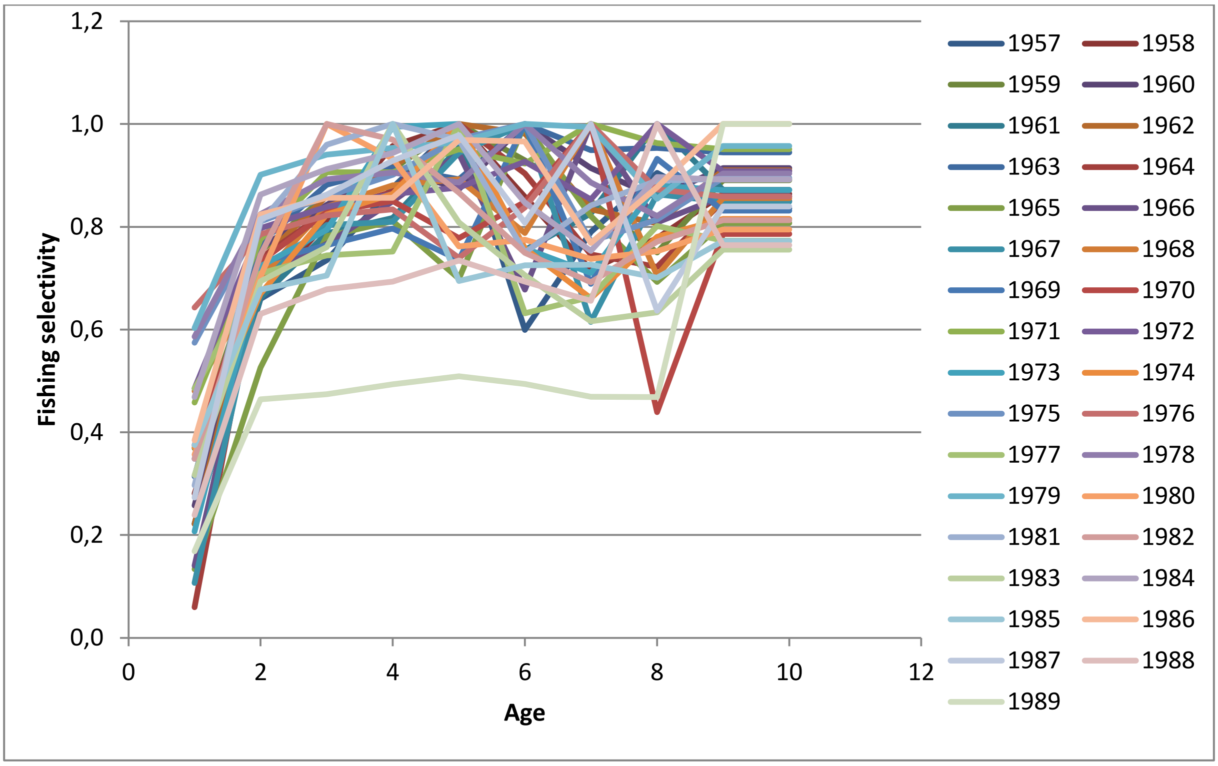
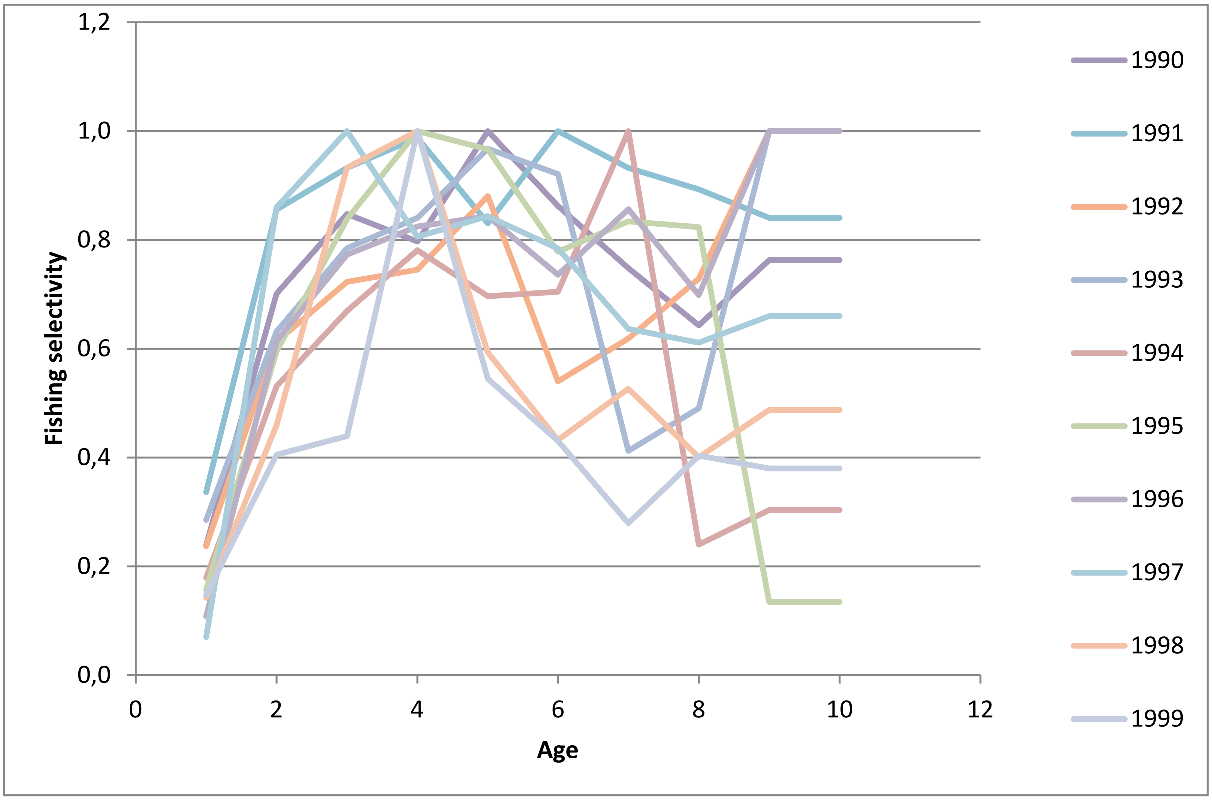


Figure A1.9. Fishing selectivities prior to the projection period from 1957 to 1989 for North Sea Plaice.



FigureA1.10. Fishing selectivities during the first decade of the projection period for North Sea Plaice.

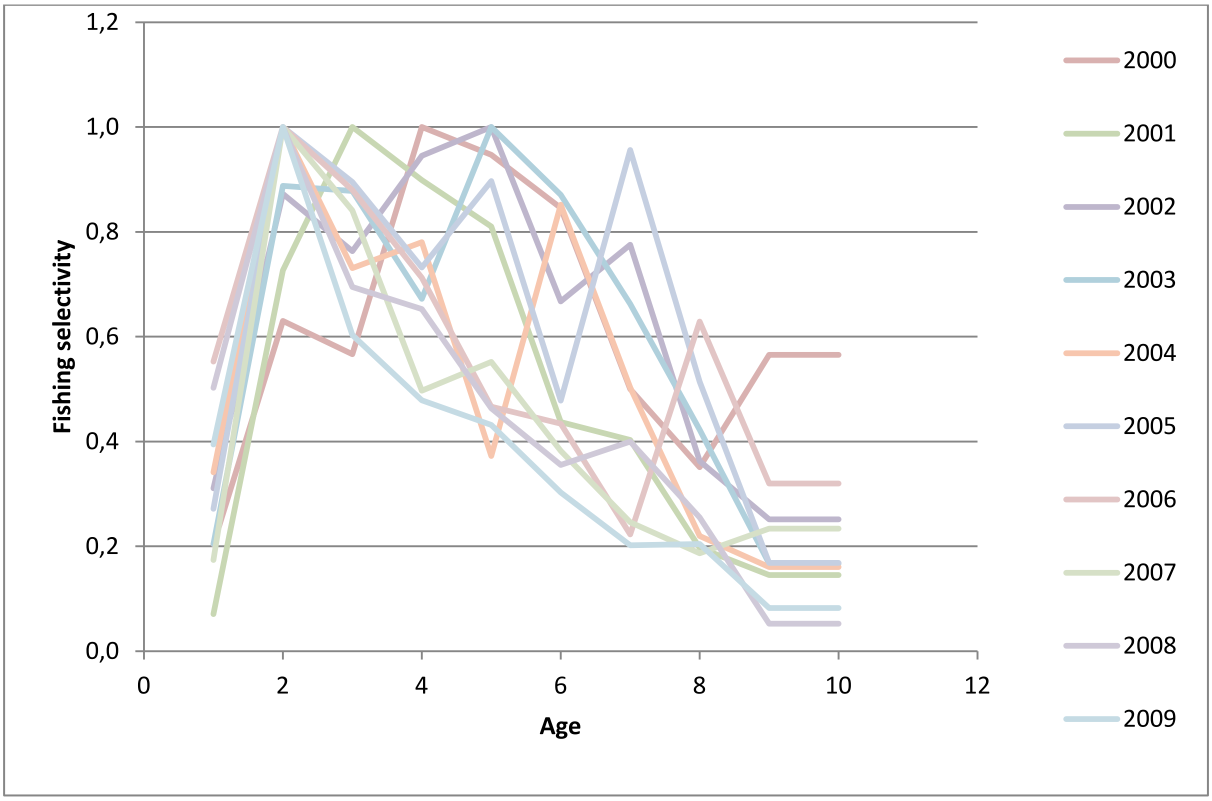


Figure A1.11. Fishing selectivities for the last decade in the projection period showing a marked decline in selectivity of older fish for North Sea Plaice.

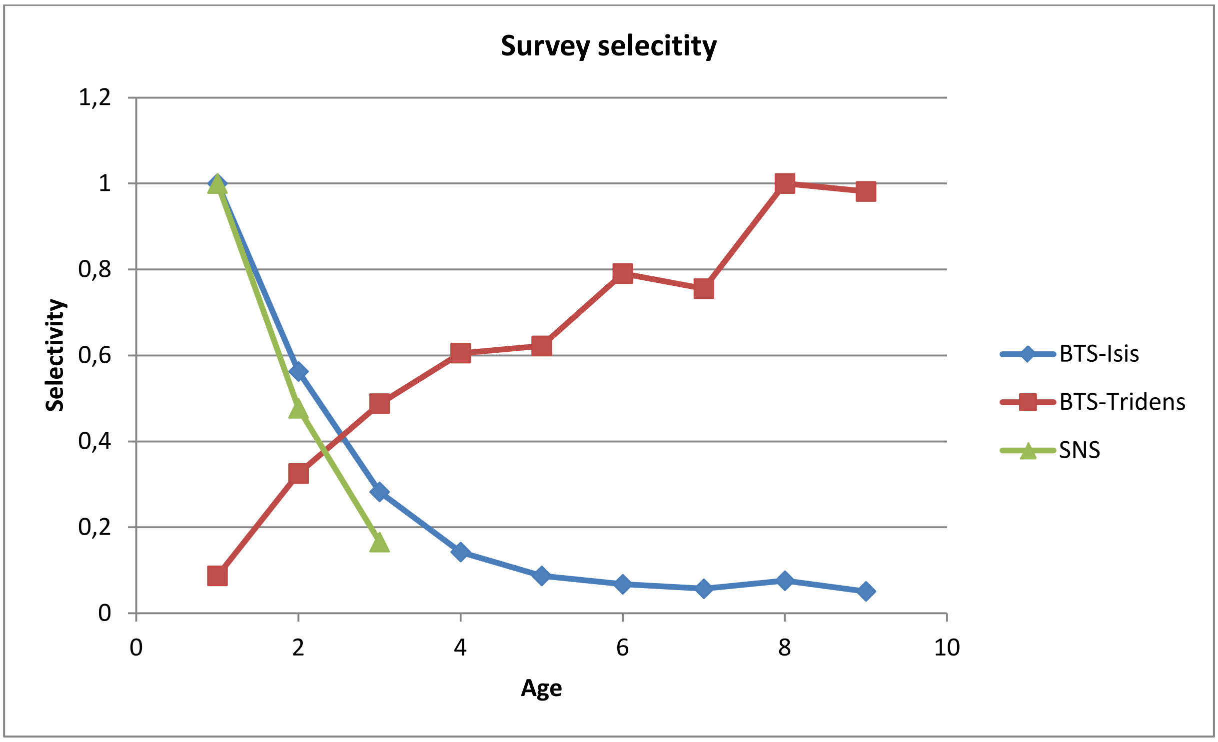


Figure A1.12. Survey selectivity vectors estimated from survey numbers-at-age as a fraction of the XSA estimated population numbers-at-age for North Sea Plaice.

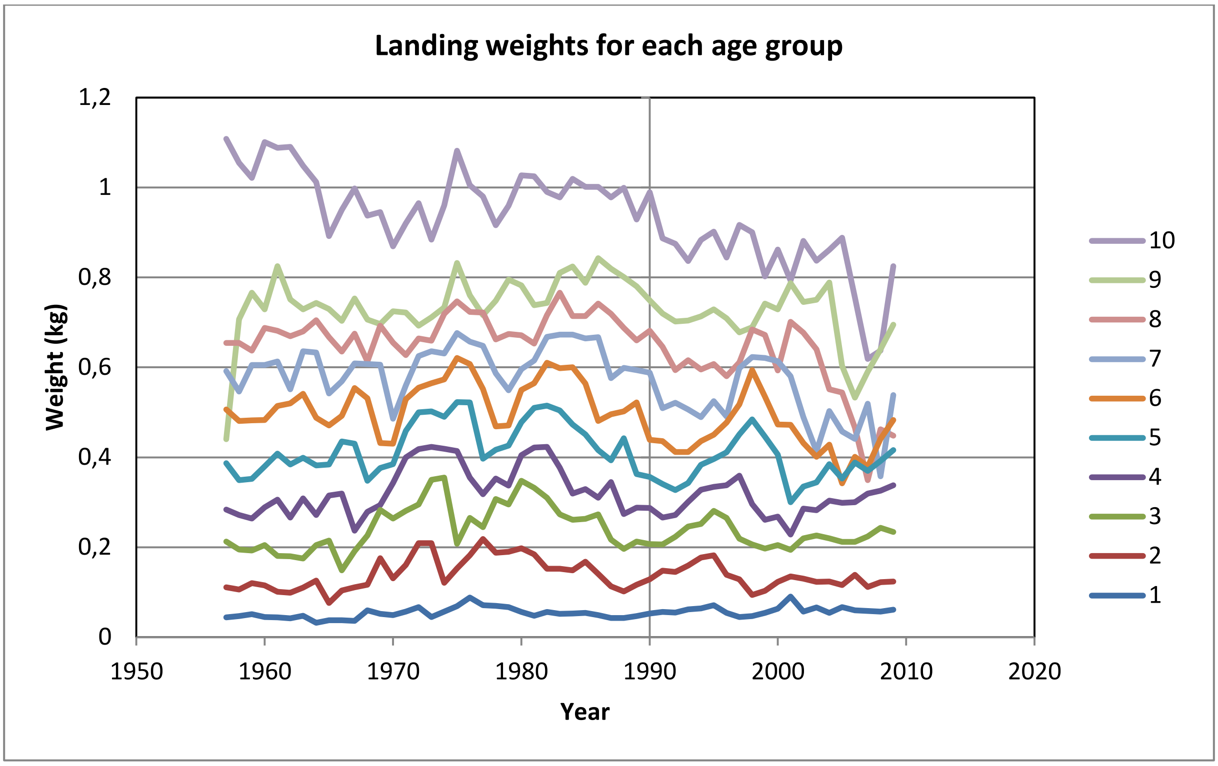


Figure A1.13. Landing weights (kg) for North Sea Plaice for each age group.

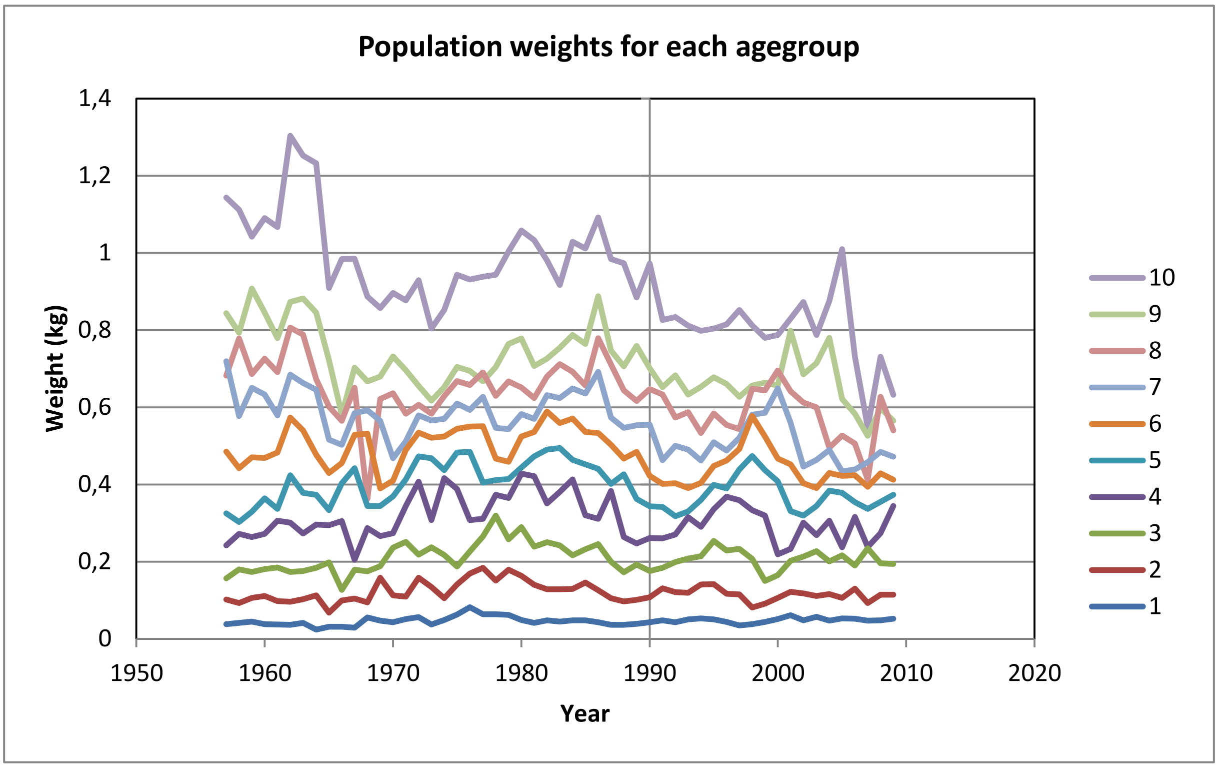


Figure A1.14. Population weights (kg) for North Sea Plaice for each age group.

Section A2: Sole in Subarea IV

The natural mortality-at-age and maturity-at-age vectors used in the XSA assessment and retrospective projections from 1990 for Sole in Subarea IV. To take into account the effect of the severe winter during 1962 to 1963, a value of 0.9 for natural mortality rate was used for 1963.

Table A2.1. Natural mortality-at-age and maturity-at-age vectors.

AGE	1	2	3	4	5	6	7	8	9	10
Natural mortality rate	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
Maturity	0.0	0.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0

The stock–recruit parameters estimated for the Beverton–Holt and 2-line functions used in the deterministic and stochastic projections.

Table A2.2. Stock–recruit parameters.

		BEVERTON–HOLT		2_LINE	
	Period	α (thousands)	β (tons)	α (thousands)	B^0 (tons)
Determinstic ("hindsight")	1957–2009	115220	8074	93345	17857
Stochastic ("forecast")	1957–1989	106895	7739	91226	22280

Year	Number of recruits	Spawning biomass	Total catch
1957	128913	60713	12067
1958	128646	64446	14287
1959	488778	66599	13832
1960	61716	71980	18620
1961	99499	113421	23566
1962	22899	111614	26877
1963	20424	106822	26164
1964	539159	36250	11342
1965	121982	28686	17043
1966	39909	83085	33340
1967	75191	81938	33439
1968	99252	68048	33179
1969	50869	51582	27559
1970	137891	44507	19685
1971	42107	39149	23652
1972	76403	43523	21086
1973	105045	34480	19309
1974	109975	33280	17989
1975	40825	35680	20773
1976	113295	37232	17326
1977	140307	30380	18003
1978	47127	34920	20280
1979	11664	42679	22598
1980	151574	32895	15807
1981	148896	22280	15403
1982	152374	31867	21579
1983	141488	39308	24927
1984	70850	42631	26839
1985	81670	39661	24248
1986	159308	32562	18201
1987	72702	28693	17368
1988	455761	38698	21590
1989	108274	33199	21805
1990	177524	89328	35120
1991	70435	77064	33513
1992	353383	76294	29341
1993	69162	54425	31491
1994	56976	74044	33002
1995	95962	58771	30467
1996	49342	37138	22651
1997	270702	29097	14901
1998	113617	20843	20868
1999	82211	41474	23475
2000	123072	38011	22641
2001	62890	30306	19944
2002	183396	30855	16945
2003	83962	24764	17920
2004	44153	36962	18757
2005	48196	31460	16355
2006	216019	23789	12594
2007	55007	17857	14635
2008	81516	37490	14071
2009	102743	34414	13952

Table A2.3. Number of recruits and spawning biomass estimates from the 2010 XSA assessment, with total annual catches for Sole in Subarea IV.

YEAR	1	2	3	4	5	6	7	8	9	10
1957	128913	72455	89309	59106	17319	15058	27046	11837	2500	46183
1958	128646	116645	64214	71157	41456	12092	10843	18272	9062	34616
1959	488778	116404	103781	50075	50907	28474	7627	6950	12311	29191
1960	61716	442265	101846	82467	35416	37526	20278	5754	4362	29304
1961	99499	55843	388723	78710	58640	23192	25996	13739	3691	22703
1962	22899	90030	49617	304373	53013	41261	16519	19770	8361	18195
1963	20424	20719	79946	38988	219104	33371	27307	10356	13977	17730
1964	539159	8304	7993	27187	10396	59622	8154	6857	2666	7985
1965	121982	487799	7366	5222	19166	5784	37457	4405	4483	6525
1966	39909	110374	396576	5629	3204	12584	2872	22002	2504	6396
1967	75191	36111	88191	231736	4152	1776	7877	1891	13893	5680
1968	99252	68036	29169	55369	128708	1898	1097	5302	988	10948
1969	50869	88820	45250	13175	26344	70258	1278	760	3234	7082
1970	137891	45652	57613	20539	6855	12054	39659	841	455	5724
1971	42107	123534	35467	27405	10751	4505	7833	24508	527	3786
1972	76403	37700	80036	18370	12662	5462	2705	4874	15314	2407
1973	105045	68792	26889	37454	9892	6734	3453	1950	3238	10857
1974	109975	94380	50614	12171	18492	5122	3883	2179	1037	7712
1975	40825	99414	70768	25308	5793	10016	2806	2006	1346	4697
1976	113295	36689	68119	36890	11754	3256	5419	1788	952	3297
1977	140307	101523	29828	35034	20050	6051	2027	3088	1095	2039
1978	47127	125294	70623	15505	17141	11065	3783	1531	1736	2139
1979	11664	42617	89560	36039	8200	9194	5969	1770	804	2135
1980	151574	10546	30781	41875	17332	4570	5255	3739	825	1821
1981	148896	136544	8392	15951	20977	8742	2755	2665	2043	1275
1982	152374	134324	95758	4493	7902	11158	4428	1592	1568	1819
1983	141488	135343	96396	43130	2313	3788	5541	2410	855	1822
1984	70850	127653	89734	47855	18870	1499	2116	3159	1250	1272
1985	81670	63926	86270	39406	21847	8712	652	1103	1866	1277
1986	159308	73741	42036	36955	16359	10823	4501	391	637	1824
1987	72702	143792	57817	20440	16600	7433	4547	1913	249	1187
1988	455761	65694	102472	31344	10030	8826	3759	2624	872	855
1989	108274	412380	46868	47840	13872	4908	4293	1926	1527	575
1990	177524	97859	329012	25036	21738	8154	2869	2518	1127	1302
1991	70435	159810	77190	198343	13363	10924	4192	1604	1139	1062
1992	353383	63618	132081	45664	105355	5695	6388	2142	767	649
1993	69162	318822	51065	77298	25905	58879	2797	2888	1233	581
1994	56976	62529	240492	30255	40065	10247	30413	1114	1487	997
1995	95962	50871	49155	134466	14487	18427	3840	16760	567	792
1996	49342	82263	33885	28476	56443	7108	9723	1576	9444	557
1997	270702	44482	56528	15260	9629	25144	2762	4251	529	5718
1998	113617	243429	34497	28652	6840	3854	10582	1356	1634	1916
1999	82211	102573	166378	16812	11719	2864	1641	5040	472	1081
2000	123072	74114	77819	81610	7424	4791	1438	853	2673	369
2001	62890	109124	52724	39272	33128	3590	1979	537	358	1745
2002	183396	56064	74154	27165	16628	14061	1891	997	225	965
2003	83962	164940	40215	35848	12889	7231	6590	1071	335	629
2004	44153	74975	118490	19743	17131	6126	2828	3651	589	521
2005	48196	39460	53619	61584	8805	8408	3484	1702	2390	307
2006	216019	42510	28650	26504	27640	3913	3852	1703	954	1597
2007	55007	188980	29247	16307	15098	14889	2045	2027	871	1385
2008	81516	49471	133054	16108	8623	8200	8228	1067	1141	807
2009	102743	71932	38866	85491	8984	5072	5188	5008	548	981

Table A2.4. Population numbers-at-age for Sole in Subarea IV taken from 2010 ICES XSA assessment, but with plusgroup adjusted as described in text.

YEAR	1	2	3	4	5	6	7	8	9	10
1957	0.000	0.021	0.127	0.255	0.259	0.228	0.292	0.167	0.241	0.241
1958	0.000	0.017	0.149	0.235	0.276	0.361	0.345	0.295	0.303	0.303
1959	0.000	0.034	0.130	0.246	0.205	0.239	0.182	0.366	0.248	0.248
1960	0.000	0.029	0.158	0.241	0.323	0.267	0.289	0.344	0.294	0.294
1961	0.000	0.018	0.145	0.295	0.252	0.239	0.174	0.397	0.272	0.272
1962	0.000	0.019	0.141	0.229	0.363	0.313	0.367	0.247	0.304	0.304
1963	0.000	0.053	0.179	0.422	0.402	0.509	0.482	0.457	0.479	0.479
1964	0.000	0.020	0.326	0.250	0.486	0.365	0.516	0.325	0.390	0.390
1965	0.000	0.107	0.169	0.388	0.321	0.600	0.432	0.465	0.443	0.443
1966	0.000	0.124	0.437	0.204	0.490	0.368	0.318	0.360	0.349	0.349
1967	0.000	0.114	0.365	0.488	0.683	0.382	0.296	0.549	0.481	0.481
1968	0.011	0.308	0.695	0.643	0.505	0.296	0.268	0.394	0.422	0.422
1969	0.008	0.333	0.690	0.553	0.682	0.472	0.318	0.412	0.489	0.489
1970	0.010	0.152	0.643	0.547	0.320	0.331	0.381	0.367	0.390	0.390
1971	0.011	0.334	0.558	0.672	0.577	0.410	0.374	0.370	0.483	0.483
1972	0.005	0.238	0.659	0.519	0.531	0.358	0.227	0.309	0.390	0.390
1973	0.007	0.207	0.693	0.606	0.558	0.451	0.360	0.532	0.503	0.503
1974	0.001	0.188	0.593	0.642	0.513	0.502	0.561	0.382	0.522	0.522
1975	0.007	0.278	0.551	0.667	0.476	0.514	0.351	0.645	0.506	0.506
1976	0.010	0.107	0.565	0.510	0.564	0.374	0.463	0.391	0.634	0.634
1977	0.013	0.263	0.554	0.615	0.494	0.370	0.181	0.476	0.282	0.282
1978	0.001	0.236	0.573	0.537	0.523	0.517	0.660	0.544	0.496	0.496
1979	0.001	0.225	0.660	0.632	0.485	0.459	0.368	0.663	0.379	0.379
1980	0.004	0.128	0.557	0.591	0.584	0.406	0.579	0.504	0.630	0.630
1981	0.003	0.255	0.525	0.602	0.531	0.580	0.449	0.430	0.501	0.501
1982	0.019	0.232	0.698	0.564	0.635	0.600	0.508	0.521	0.520	0.520
1983	0.003	0.311	0.600	0.727	0.334	0.482	0.462	0.556	0.644	0.644
1984	0.003	0.292	0.723	0.684	0.673	0.733	0.552	0.426	0.581	0.581
1985	0.002	0.319	0.748	0.779	0.602	0.561	0.411	0.448	0.444	0.444
1986	0.002	0.143	0.621	0.700	0.689	0.767	0.756	0.351	0.629	0.629
1987	0.001	0.239	0.512	0.612	0.532	0.582	0.450	0.686	0.419	0.419
1988	0.000	0.238	0.662	0.715	0.615	0.621	0.569	0.442	0.999	0.999
1989	0.001	0.126	0.527	0.689	0.431	0.437	0.434	0.436	0.379	0.379
1990	0.005	0.137	0.406	0.528	0.588	0.565	0.482	0.694	0.727	0.727
1991	0.002	0.091	0.425	0.533	0.753	0.436	0.572	0.637	1.121	1.121
1992	0.003	0.120	0.436	0.467	0.482	0.611	0.694	0.452	0.791	0.791
1993	0.001	0.182	0.423	0.557	0.827	0.561	0.820	0.564	0.499	0.499
1994	0.013	0.141	0.481	0.636	0.677	0.882	0.496	0.576	1.043	1.043
1995	0.054	0.306	0.446	0.768	0.612	0.539	0.790	0.474	0.792	0.792
1996	0.004	0.275	0.698	0.984	0.709	0.845	0.727	0.991	0.459	0.459
1997	0.006	0.154	0.580	0.702	0.816	0.765	0.611	0.856	1.082	1.082
1998	0.002	0.281	0.619	0.794	0.771	0.754	0.642	0.955	1.089	1.089
1999	0.004	0.176	0.612	0.717	0.794	0.589	0.554	0.534	1.336	1.336
2000	0.020	0.241	0.584	0.802	0.627	0.784	0.886	0.768	0.456	0.456
2001	0.015	0.286	0.563	0.759	0.757	0.541	0.585	0.769	0.679	0.679
2002	0.006	0.232	0.627	0.646	0.733	0.658	0.469	0.992	0.537	0.537
2003	0.013	0.231	0.611	0.638	0.644	0.839	0.490	0.499	0.515	0.515
2004	0.012	0.235	0.554	0.707	0.612	0.464	0.408	0.324	1.187	1.187
2005	0.026	0.220	0.605	0.701	0.711	0.681	0.616	0.479	0.424	0.424
2006	0.034	0.274	0.464	0.463	0.519	0.549	0.542	0.570	0.511	0.511
2007	0.006	0.251	0.496	0.537	0.510	0.493	0.551	0.475	0.928	0.928
2008	0.025	0.141	0.342	0.484	0.431	0.358	0.397	0.567	0.586	0.586
2009	0.017	0.164	0.341	0.391	0.479	0.415	0.385	0.384	0.899	0.899

Table A2.5. Fishing mortality-at-age for Sole in Subarea IV taken from the 2010 ICES XSA assessment, but with plusgroup adjusted as described in text.

YEAR	BTS-ISIS	SNS
1957		
1958		
1959		
1960		
1961		
1962		
1963		
1964		
1965		
1966		
1967		
1968		
1969		
1970		293.271
1971		326.698
1972		129.932
1973		388.106
1974		149.056
1975		188.455
1976		113.629
1977		278.598
1978		312.679
1979		151.525
1980		222.524
1981		396.531
1982		487.724
1983		323.870
1984		355.336
1985	2.709	315.860
1986	2.003	267.961
1987	3.190	260.693
1988	6.819	716.098
1989	11.912	863.145
1990	11.015	523.502
1991	7.270	667.303
1992	11.310	698.119
1993	10.028	579.423
1994	5.658	268.420
1995	6.043	226.161
1996	3.088	75.031
1997	10.290	579.225
1998	5.459	690.122
1999	6.940	297.214
2000	2.854	156.842
2001	2.810	170.643
2002	2.772	470.770
2003	3.117	
2004	1.898	247.183
2005	1.664	76.286
2006	1.637	166.041
2007	3.942	195.469
2008	5.085	167.186
2009	3.350	180.814

Table A2.6. Age-aggregated survey biomass indices for Sole in Subarea IV.

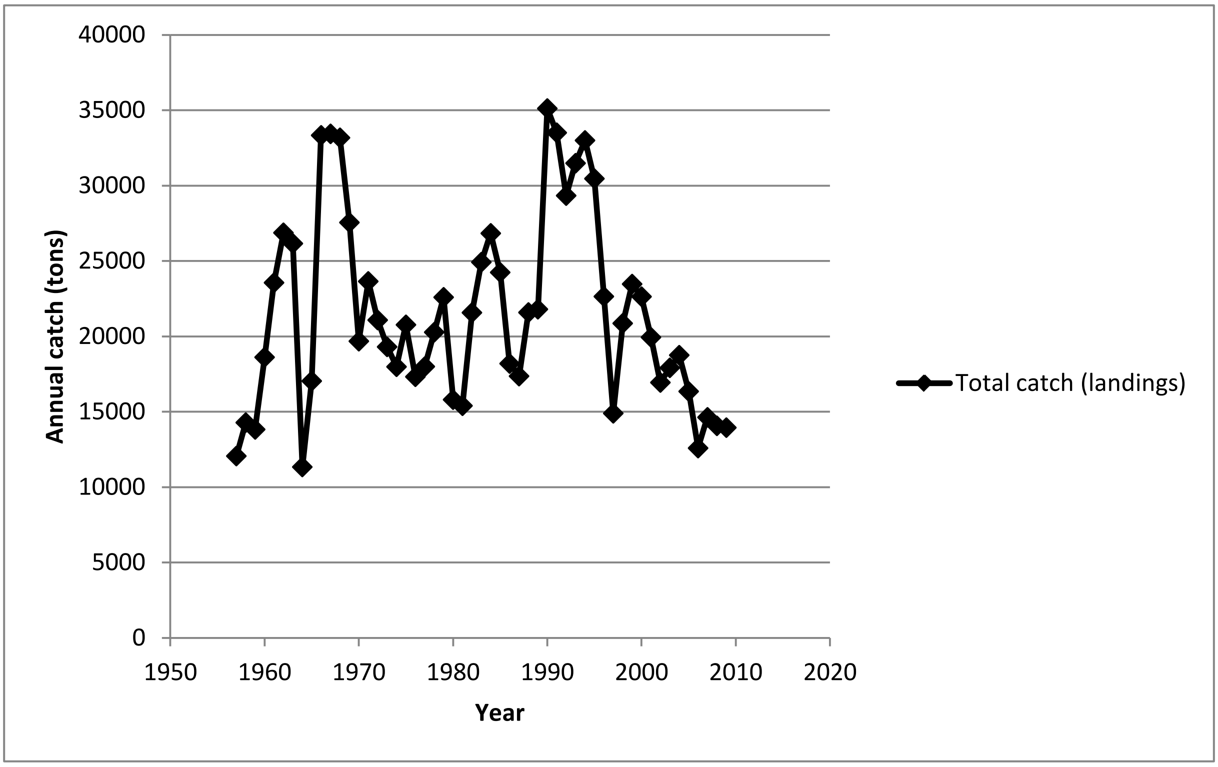


Figure A2.1. Total annual landings of Sole in Subarea IV in tons.

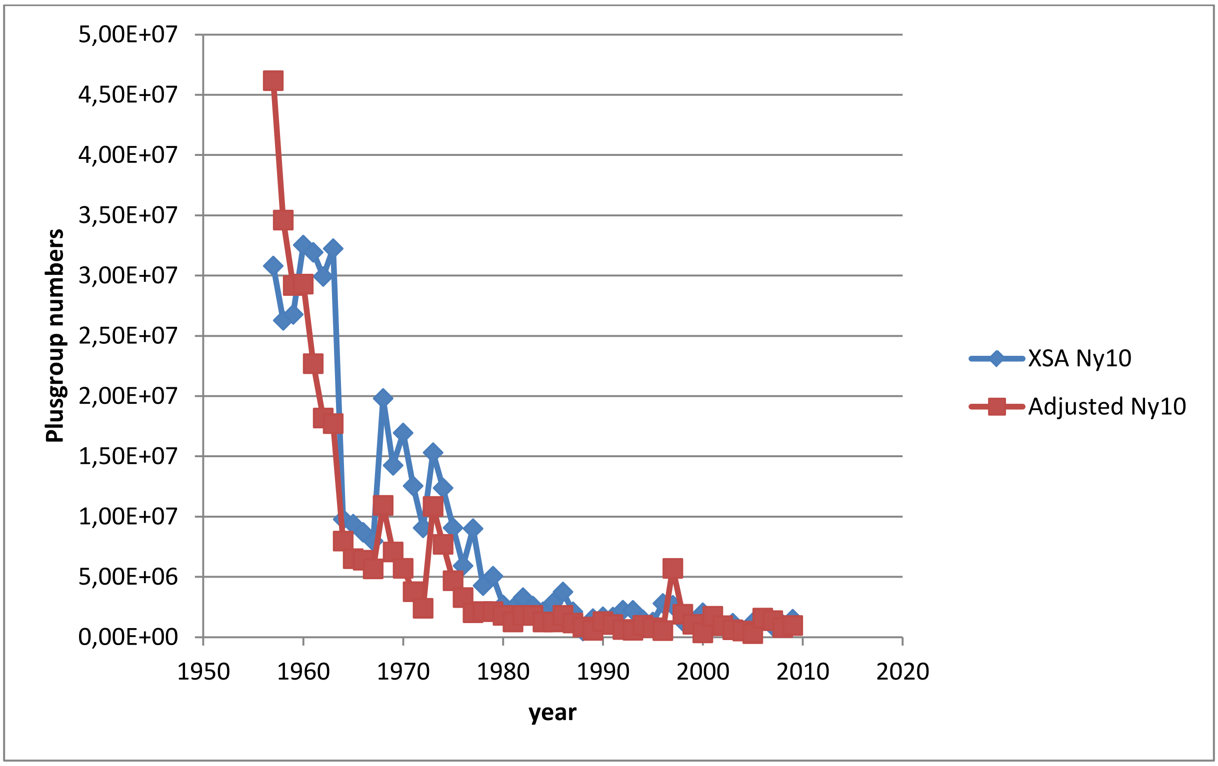


Figure A2.2. Adjusted plusgroup population numbers compared to the XSA estimates for Sole in Subarea IV.

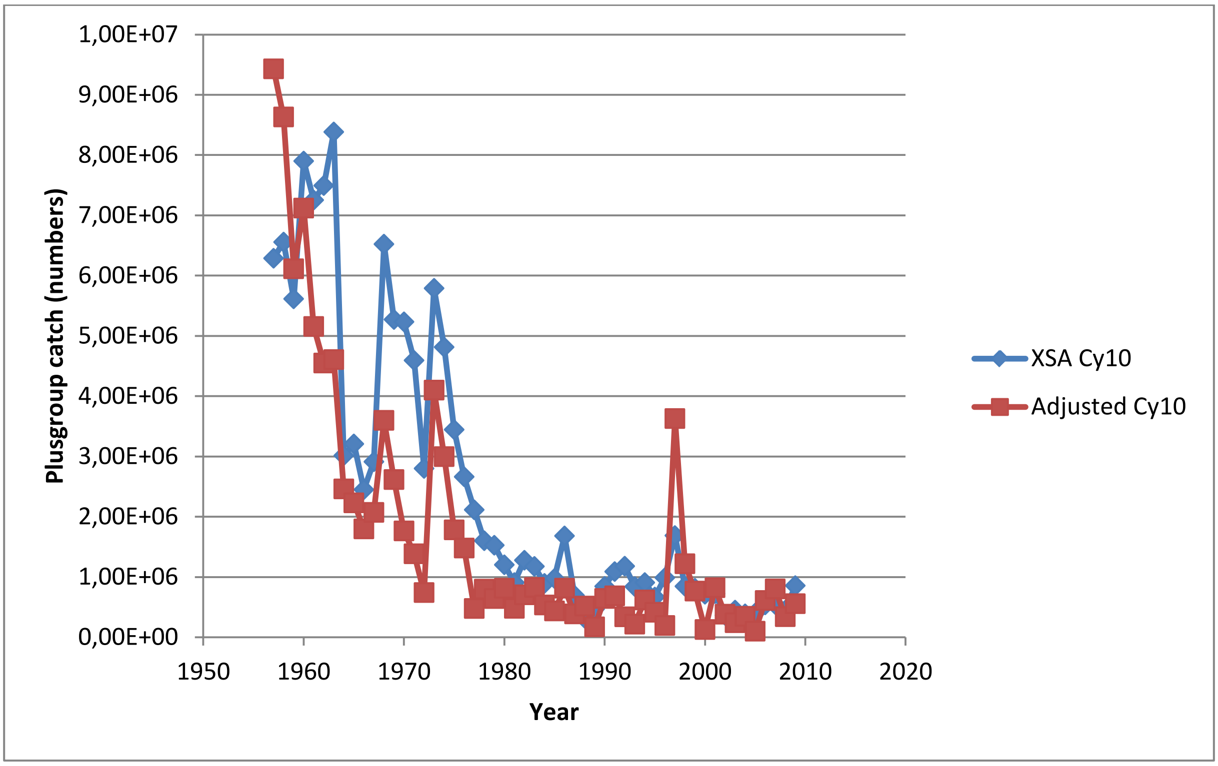


Figure A2.3. Adjusted plusgroup catch compared to observed catch for Sole in Subarea IV.

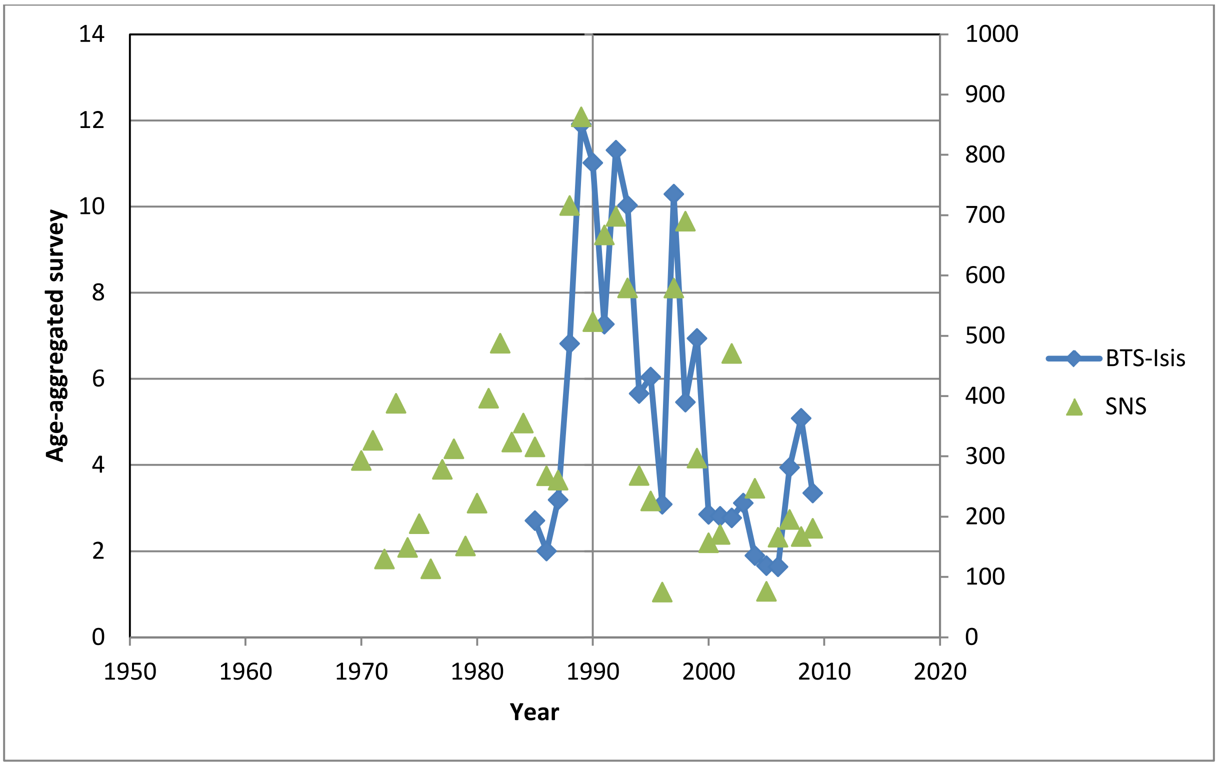


Figure A2.4. Age-aggregated survey series for Sole in Subarea IV.

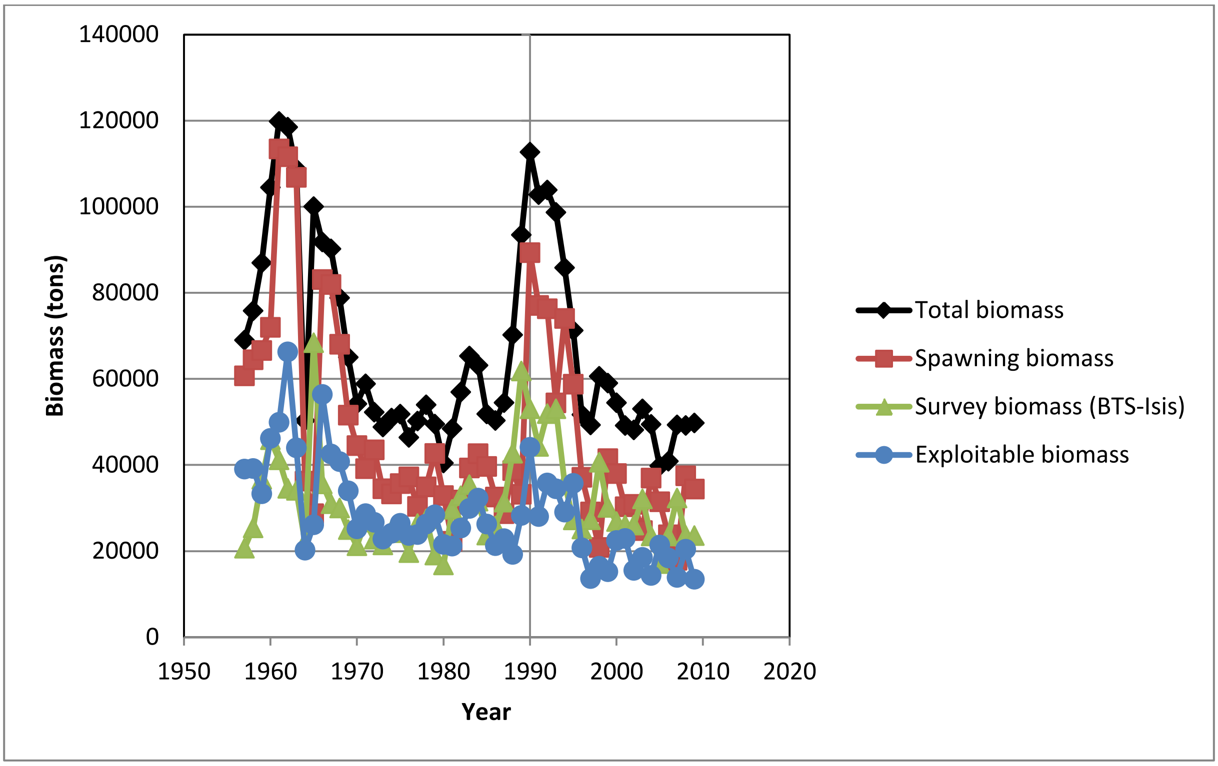


Figure A2.5. Trajectories for various biomass components from the XSA assessment for Sole in Subarea IV with plusgroup adjusted as detailed in the text.

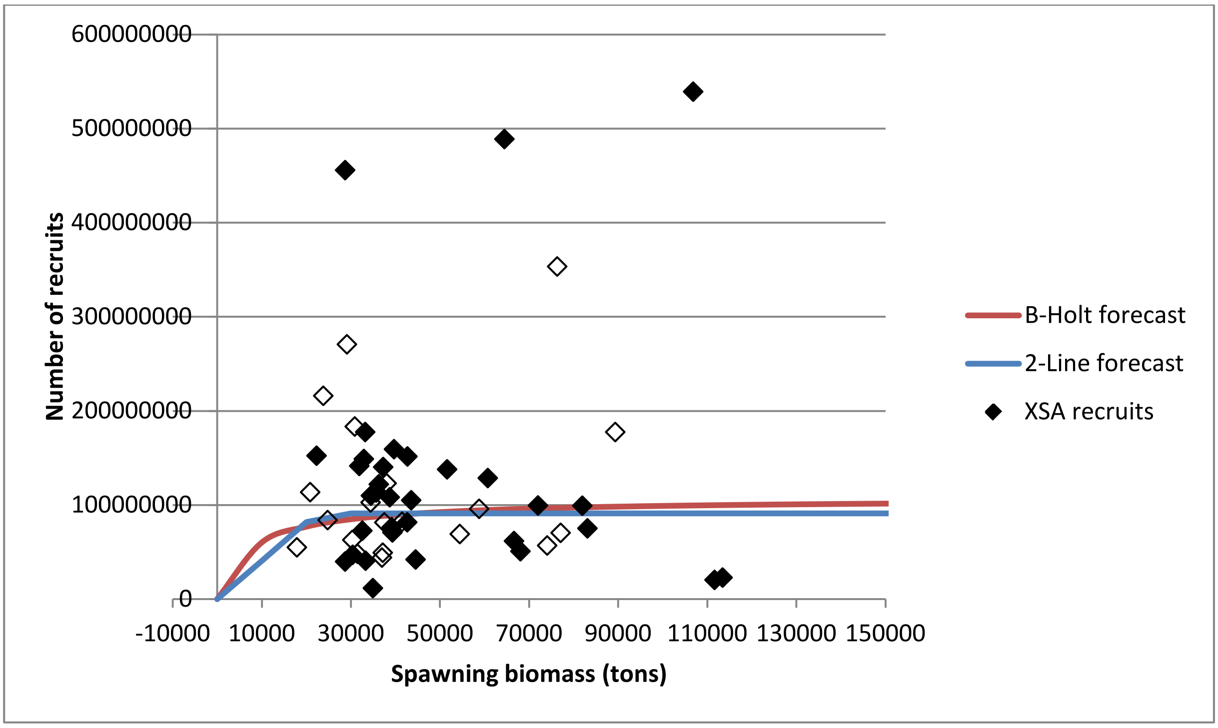


Figure A2.6. Number of recruits (1-yr-olds) estimated in the 2010 XSA assessment for Sole in Subarea IV (diamonds) and number of recruits in terms of a Beverton–Holt stock–recruitment curve when fixing *h*=0.9 and a 2-line stock–recruit relationship fitted to data from 1957 to 1989 (forecast) . Recruitments from 1990 onwards are shown by open diamonds.

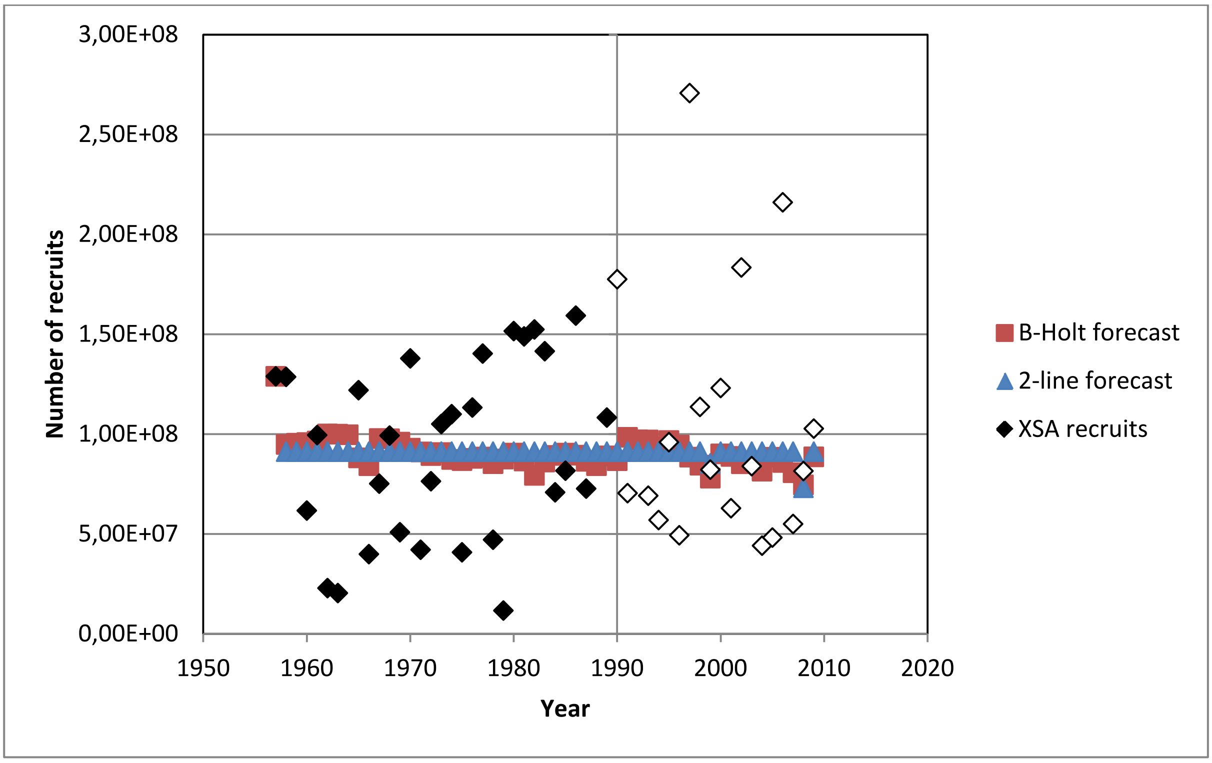


Figure A2.7. Annual number of recruits (1-yr-olds) estimated in the 2010 XSA assessment for Sole in Subarea IV (diamonds) compared to the number of recruits in terms of a Beverton–Holt stock–recruitment curve when fixing *h=0.9* (squares) and a 2-line stock–recruit relationship.

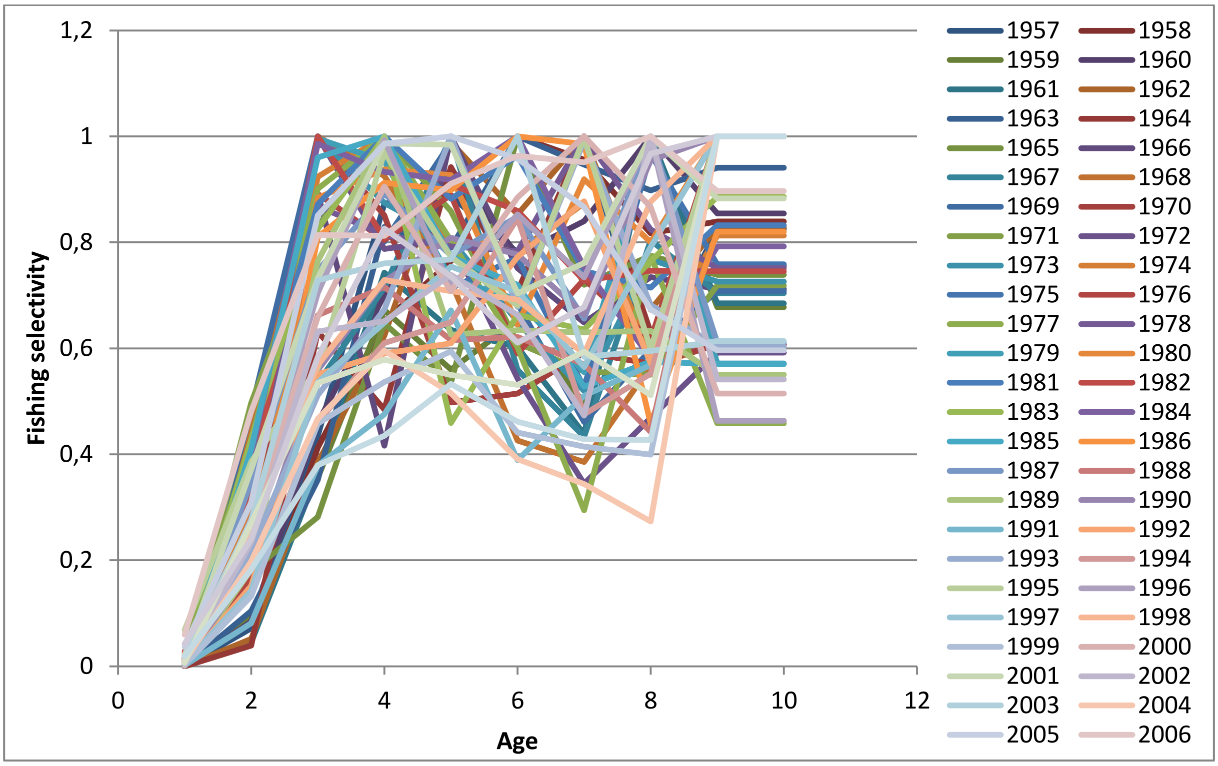


Figure A2.8. Fishing selectivities-at-age over the assessment period from 1957 to 2009 for Sole (Subarea IV).

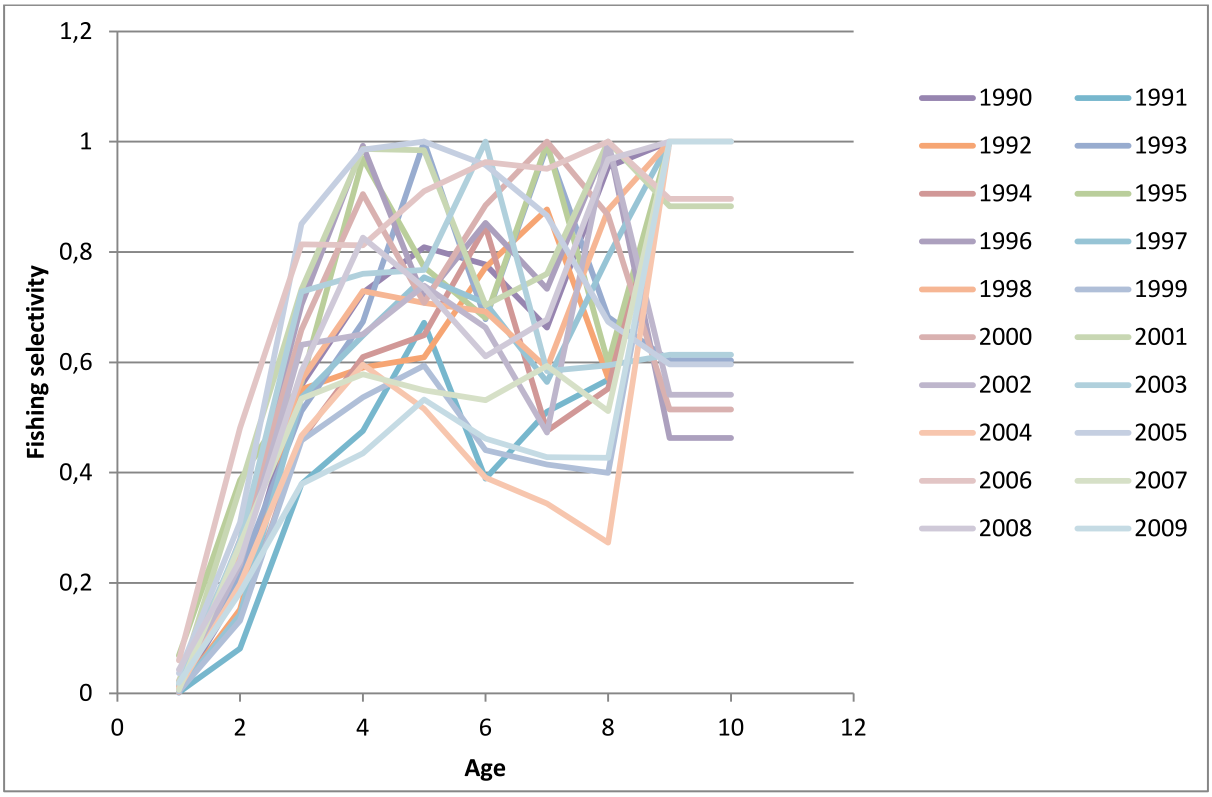


Figure A2.9. Fishing selectivities for Sole in Subarea IV during the projection period.

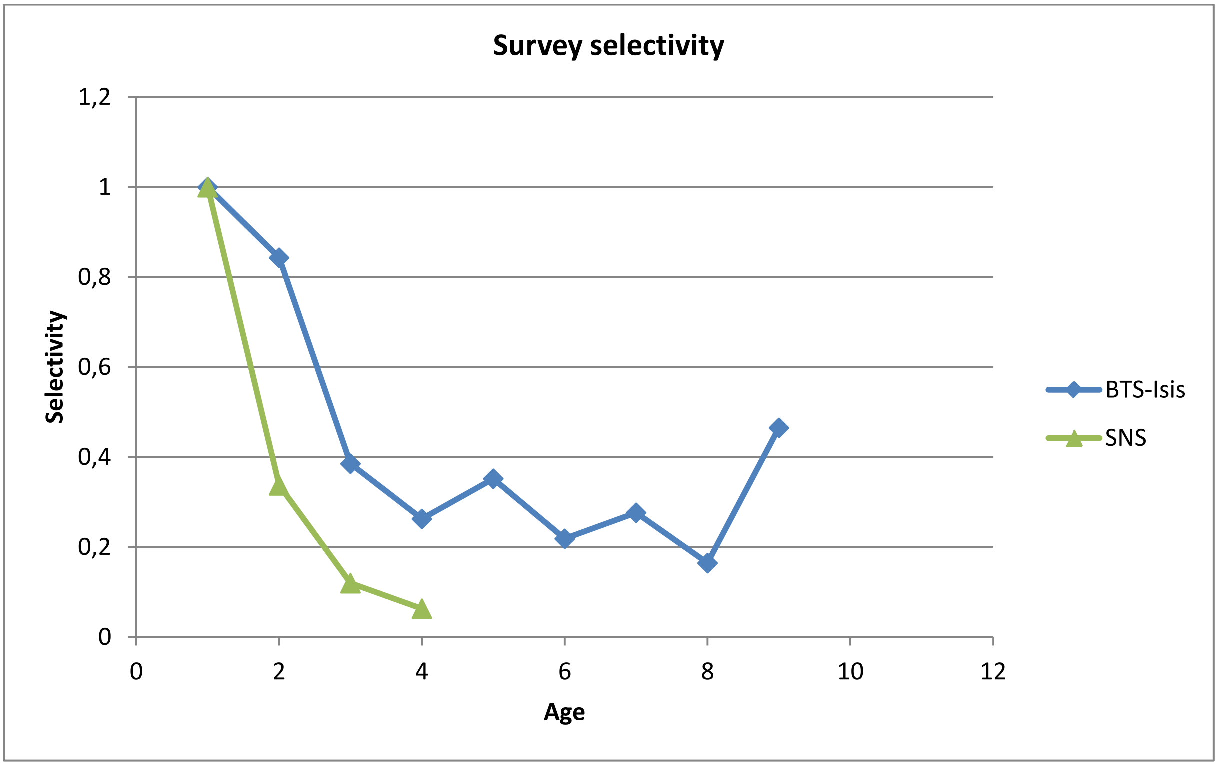


Figure A2.10. Survey selectivity vectors estimated from survey numbers-at-age as a fraction of the adjusted XSA estimated population numbers-at-age for Sole in Subarea IV.

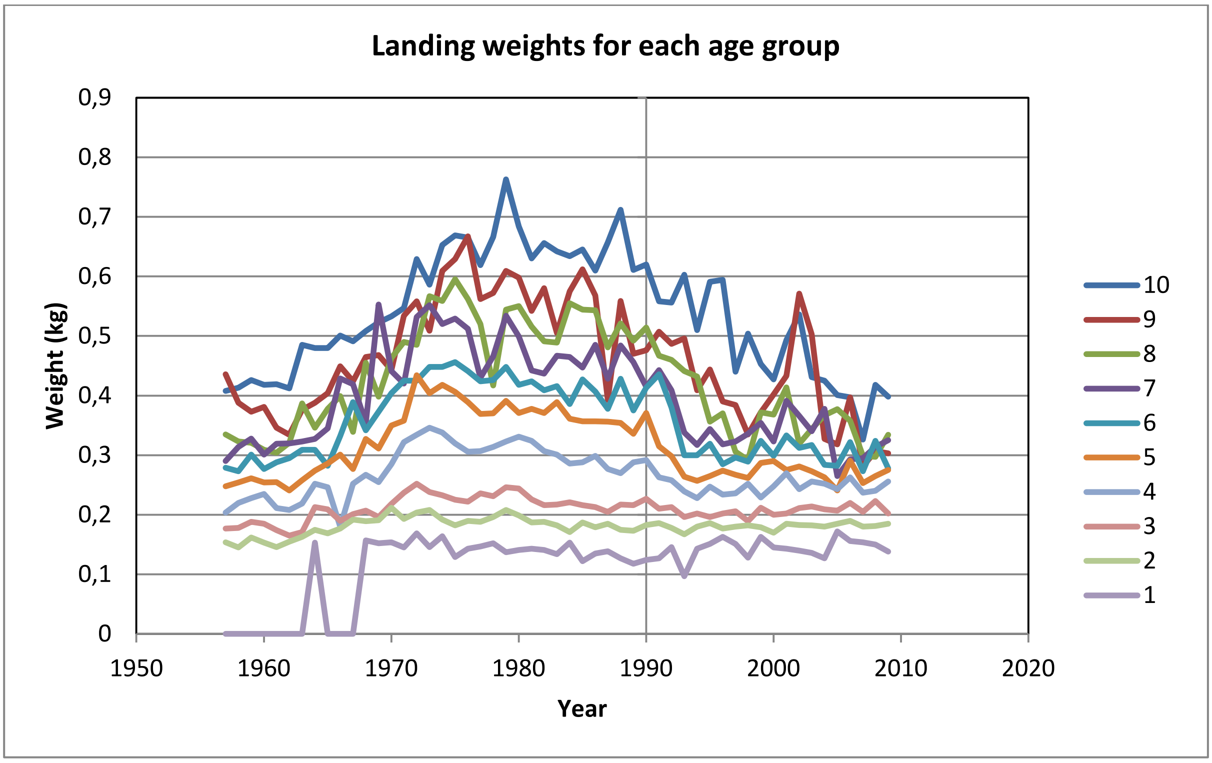


Figure A2.11. Landing weights (kg) for Sole in Subarea IV for each age group.

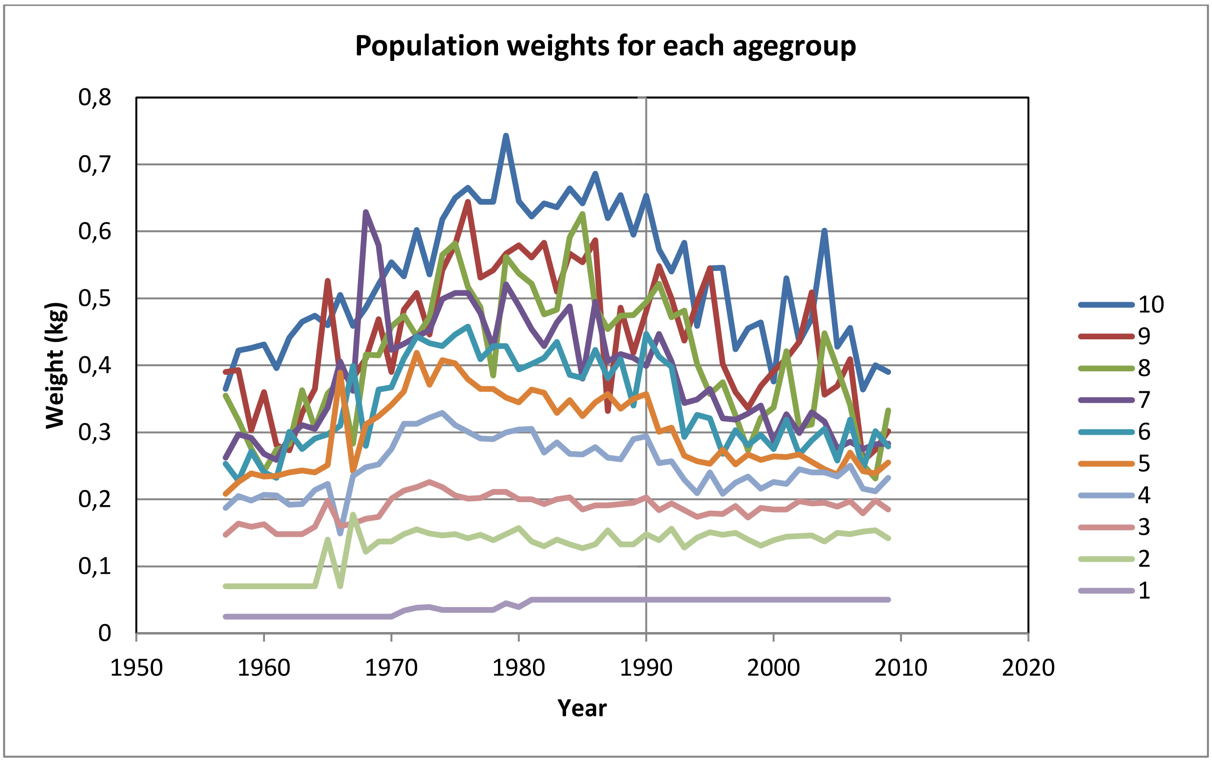


Figure A2.12. Population weights (kg) for Sole in Subarea IV for each age group.

Working Document 8

Correcting for measurement error bias when fitting stock–recruit models and estimating MSY reference points.

Noel G. Cadigan and Joanne Morgan

Fisheries and Oceans Canada, Northwest Atlantic Fisheries Center, 80 East White Hills Road, St John's, NL, Canada A1C 5X1
709 772 5028 (Ph), 709 772 4188 (Fax)

Abstract

Understanding the relationship between spawning stock size and subsequent reproduction and recruitment of juveniles to a fishery is a fundamental component of sustainable fisheries management. Literature results suggest that the effect of parental stock size measurement errors (SME's) when fitting spawner-recruitment models is to overestimate the slope at the origin and underestimate the maximum recruitment. This will often lead to poor estimates of MSY reference points, which have been adopted by many national and international fisheries management agencies to guide fisheries management.

We present methods to account for SME when fitting spawner-recruitment models and estimating Fmsy and Bmsy, and apply these methods to two case studies of cod and American plaice in NAFO Subdividion 3Ps, located off the south coast on Newfoundland, Canada. If the SME is large then estimates of MSY reference points can be quite different from the results obtained assuming no ME or only ME in the population numbers-at-age component of parental stock size. Fmsy may be considerably lower, and Bmsy considerably greater, than the no-ME results. However, the SME variance is highly confounded with ME and process error in recruitment, and additional data are likely to be required to estimate the SME variance reliably.

Introduction

Understanding the relationship between parental stock size (*S*) and subsequent reproduction and recruitment (*R*) of juveniles to a fishery is widely recognized as a fundamental component of sustainable fisheries management (Quinn and Deriso, 1999). For example, stock–recruit (SR) relationships are used to project future fish population dynamics in response to proposed management actions, and to determine management reference points (Needle, 2002). Many fisheries are managed using reference points (RPs), where prescribed actions should occur when stock size or fishing mortality rates transgress reference points. Reference points are widely considered an essential part of well-managed fisheries (e.g. Hilborn and Stokes, 2010). Reliable SR models are therefore important for successful fisheries management.

Maximum sustainable yield (MSY) RPs have been adopted by many national (e.g. US, NZ) and international fisheries management agencies (e.g. IWC, ICCAT, IATTC, ICES, NAFO). The fishing mortality that maximizes long-term yield (Fmsy) is often taken to be an upper limit for management purposes, while the resulting biomass at

Fmsy (i.e. Bmsy) is a target. Although management using MSY RPs is not without criticism (e.g. Hilborn, 2010; Legović *et al.*, 2010), if these RPs are to be used then it is important to have reliable estimates of them.

Within age structured models, MSY RPs are essentially derived from long-term stock projections over a range of fishing mortalities. If the projections are deterministic then the calculation of MSY RPs is also deterministic. In this context some theory has been developed related to MSY calculations (Sissenwine and Shepherd, 1987). MSY RPs are determined by the growth and mortality process of the stock, and by the age pattern in fishing mortality. The SR relationship is a fundamental component in the stock growth process. Consequently, the SR relationship has a major impact on MSY RPs.

Parametric SR models are commonly used to compute MSY RPs, especially when MSY calculations involve inferring R outside the range of the estimated S's. A parametric SR model expresses R as an analytic function of S and a small number of unknown parameters θ that must be estimated. Two SR models commonly used are the Ricker (Ricker, 1954) and the Beverton–Holt (Beverton and Holt, 1957). These models are described in the Methods section. The SR θ parameters are estimated either as part of an analytic stock assessment model, or based on a time-series of SR estimates obtained from a stock assessment model. In the latter case, which we refer to as external SR estimation, the θ parameters are usually estimated by minimizing the log error sum of squares based on a sample of SR observations, although other estimation procedures have been advocated (e.g. Walters 1990; Michielsens and McAllister, 2004; Jiao *et al.*, 2004). We refer to fitting a SR model as part of an assessment model as internal estimation.

The log error sum of squares fitting criteria is based on the assumption that S is known with little or no error and most of the error in the SR data comes from the measurement of R and other process errors or environmental variability that affect how much R is derived from S amount of parents. This is the common regression estimation framework where the relationship between known covariates and a random response is estimated. However, in most situations S is not known without error. It has been argued that the measurement error (ME) in R is larger than the ME in S because S is usually derived as the sum of estimates of biomass-at-age times maturity-at-age from some type of cohort model, and age-specific errors in biomass estimates will tend to cancel in the sum. This is true, but there are additional sources of ME in S related to how adequately S reflects the actual reproductive potential of a stock. Trends in sex ratio and fecundity as well as possible effects from changes in reproductive potential related to changes in age composition are not accounted for in the normal calculation of S (Morgan *et al.*, 2011). Hence, the ME errors in S may actually be of equivalent or larger in magnitude compared to the ME and process errors in R. When the SR parameters are estimated internally then the ME in S caused by estimation errors in biomass-at-age may be properly adjusted for, but the other sources of ME in S still need to be accounted for.

It is well known (e.g. Carroll *et al.*, 2006) that ME can result in biased estimates of the parameters of regression models. For example, consider the common simple linear regression model for a response variable Y that is a stochastic linear function of some covariate X, $Y = \beta_o + \beta_1 X + \varepsilon_Y$, where ε_Y is a random error term that is usually assumed to be distributed as $\varepsilon_y \sim N(0, \sigma_\varepsilon^2)$. For convenience it is assumed that X is also random with mean μ_X and variance τ_X^2. If X is only observed with error, say $W = X + \varepsilon_X$, and if the regression parameters are estimated using the data $\{(y_1,w_1), (y_2,w_2), \ldots\}$, then the least squares estimates of the parameters are biased and the bias depends on the magnitude the ME's, ε_X's. If $Var(\varepsilon_X) = \sigma_X^2 = \phi\tau_X^2$ then the estimate of the slope, $\hat{\beta}_1$, has expectation $E(\hat{\beta}_1) = \lambda\beta_1$ where $\lambda = 1/(1+\phi)$. Note that $\lambda < 1$ so that the bias is attenuated towards zero. The bias depends on how large the ME variance σ_X^2 is relative to the total variability of the X's, τ_X^2. The bias is zero only if $\phi = 0$ and the bias is large when ϕ is large. A large amount of ME masks the linear relationship between Y and X. Large sample sizes do not reduce this bias.

ME bias also occurs when fitting SR models. Walters and Ludwig (1981) found that ME with the Ricker model favoured overexploitation, except if density-dependence in the data were strong, in which case ME favoured underexploitation. Kehler *et al*. (2002) found that ME led to overestimation of the slope-at-the-origin (*Sao*) of the Ricker model when most of the observed *S*'s were less than *Smax* – the *S* corresponding to maximum *R*, (*Rmax*). They found the reverse when most of the observed *S*'s were greater than *Smax*. This was basically the same conclusion as Walters and Ludwig (1981). Kope (2006) found that large ME biases led to overestimates of productivity for populations with low productivity or populations that were overfished, in which case most observations would be less than *Smax*. Cadigan (2009) presented diagnostics to describe the impact of ME on *Sao* and the *S* corresponding to 50% of *Rmax*, which we refer to as *S50%*. For the Beverton–Holt model, he concluded that ME always led to overestimation of *Sao* and underestimation of *S50%*. The ME effect for the Ricker model was the same, except when there was substantial density-dependence (i.e. decline in recruitment at large S) in the data. Overestimating *Sao* makes a stock appear more productive than it actually is.

Similar to the linear regression model, ME in SR data masks the relationship between *S* and *R*. ME reduces the apparent change in *R* as *S* increases. This may be accommodated in model fitting by attenuating towards zero the slope of the SR model within the range of the observed *S*'s. Because SR models go through the origin, this attenuation is achieved by increasing *Sao*. The other consequence of ME is underestimation of *Rmax* and S50%. This is illustrated with a simple simulated example in Figure 1. A relatively large number of SR observations (n = 100) were generated from a Beverton–Holt model using a lognormal *R* ME distribution with a coefficient of variation (CV) equal to 0.2. The *S's* were uniformly distributed over S20% - S80%. The true values of *S* were used for estimation (top panel), as well as *S* values with ME (CV equal to 0.4; bottom panel). This large amount of *S* ME was used to illustrate the bias problem. The estimated Beverton–Holt model with ME is more "knife-edged". When there was only ME in R (top panel), the estimates of *Sao* and *S50%* were 1.01 and 19 170 which were close to the true values (i.e. 1.0 and 20 000). However, when there was both ME in R and S then the estimates of *Sao* and *S50%* were 1.51 and 10 510 which were substantially different from the true values. We repeated the simulation with *n = 10 000* SR observations. As expected, the estimates of *Sao* and *S50%* with only ME in R were almost exactly equal to the true values, while the estimates with S ME were substantially biased (*Sao = 1.49; S50% = 11 096*). This demonstrates that ME bias is not reduced with large sample sizes.

A problem with this simple example is that estimates of S are highly autocorrelated, and simply adding additional error to S like above may create a time-series of S's that have low likelihood or plausibility given our other information on S from surveys, catches, etc. ME's in S related to trends in sex ratio's or changes in the age composition of S will likely be autocorrelated. We will address this problem in this paper.

For linear models it is well known that not all of the variance parameters are identified. In particular, the ME variance parameter σ_S^2 cannot be estimated without additional data (e.g. replicates) or assumptions. However, for non-linear models these parameters are sometimes identifiable (see Section 8.1.2 in Carroll *et al.*, 2006). This was observed for the Beverton–Holt and Ricker SR models. In exploratory simulations with large sample sizes, the profile likelihood for the ME variance in *S* was maximized at the correct value. However, for realistic sample sizes the likelihood was sometimes maximized at very different values of ME variance. Hence, there are parameter identification problems for practical sample sizes. We explore this issue using sensitivity analyses.

Carroll *et al*. (2006) describe several methods to reduce covariate ME bias in non-linear regression models. In regression calibration the unobserved covariate *X* in the regression model is estimated based on auxiliary information (i.e. internal validation data, replicate measurements of W, etc). Such information will usually not be available when fitting SR models. Also, Carroll *et al*. (2006) concluded that this approach can be poor for highly non-linear models. The simulation-extrapolation method involves adding additional covariate ME to the data in a resampling strategy and then establishing the bias trend with increasing ME. This trend is extrapolated to the origin to estimate the ME bias and to bias-correct parameter estimates. It requires either knowing the ME variance or having data (i.e. replicates) to estimate it. The extrapolation step can be error-prone and is more successfully performed when one has a theoretical understanding of the form of the covariate ME bias as a function of the ME variance. This understanding may be specific to the functional form of the non-linear model. Carroll *et al*. (2006) discussed a corrected score function approach which can be more generally implemented using a Monte Carlo sampling approach. They also discussed the likelihood approach which is applicable to a wide range of models and usually gives more efficient estimates than other approaches, albeit at a cost of additional assumptions and less robustness. This approach can be used directly for model comparisons and profile likelihood confidence intervals for the parameters of non-linear models. No specialized statistical theory is required for inferences. The main difficulty with this approach is computing the likelihood function which involves integration over the unobserved X variables. Fortunately, this is easy in ADMB (ADMB Project 2009), and this is the approach we pursue.

ME bias may lead to overestimation of Fmsy and underestimation of Bmsy. If these RP's are used in a fisheries management framework then the biases will lead to a smaller "caution/critical" zone which could possibly lead to overexploitation of stocks. In this paper the impact and magnitude of ME in S on MSY RP's derived from Beverton–Holt and Ricker SR models is quantified. Methods are illustrated using two example stocks, cod and American plaice in NAFO Subdivision 3Ps.

Material and methods

SR models and RP's

Let $\mu(s) = E(R|S{=}s)$ denote the SR model which gives the expected value of the recruitment random variable (R) as a function of stock size (S). The Beverton–Holt SR model is $\mu(s) = \alpha s/(\beta+s)$. It is straight-forward to show that $Rmax = \alpha$, $S50\% = \beta$, and Sao = α/β. The Ricker model is $\mu(s) = \alpha s \exp(-\beta s)$. $Rmax = \alpha/\beta \exp(1)$ and Sao = α; however, a closed form solution for $S50\%$ does not exist and it must be found numerically. The Ricker model is commonly used for SR relationships in which a reduction in recruitment at large stock sizes is expected to occur because of density-dependent effects.

Deterministic MSY RP's were obtained for each SR model using the approach of Sissenwine and Shepherd (1987). This requires additional information on weights, maturities, natural mortality, and fishery selectivity. These are described separately for each example.

External SR estimation methods

The data available are a paired time-series $\{(R_1,S_1),\ldots,(R_n,S_n)\}$ obtained from fitting a stock assessment model or directly from a survey. The length of the time-series is denoted by n. Both R and S are assumed to be measured with error. Let S^T and R^T denote the true values of S and R,

$$R = R^T \exp(\varepsilon_{RME} - \sigma^2_{RME}/2),\ \ \varepsilon_{RME} \sim N(0,\sigma^2_{RME}) \qquad (1)$$

and

$$S = S^T \exp(\varepsilon_{SME} - \sigma^2_{SME}/2),\ \ \varepsilon_{SME} \sim N(0,\sigma^2_{SME}) \qquad (2)$$

Note that errors in Eq.s (1) and (2) are adjusted so that $E(R) = R^T$ and $E(S) = S^T$. We assume that the R and S ME's are independent of each other, and independent over time. In fact there will be some time-series correlations because current S is a function of previous R's. Also, S ME's related to trends in sex ratio's or changes in the age composition of S will likely be autocorrelated. We account for this in the next section.

Let $\mathbf{R}$ and $\mathbf{R}^T$ be $n\times 1$ vectors of R estimates and true values, and define $\mathbf{S}$ and $\mathbf{S}^T$ similarly. The likelihood of the "data" (i.e. R and S estimates) is based on $\Pr(\mathbf{R} = \mathbf{r}, \mathbf{S} = \mathbf{s})$. Given values of $\mathbf{R}^T$ and $\mathbf{S}^T$, the conditional probability of the data, $\Pr(\mathbf{R} = \mathbf{r}, \mathbf{S} = \mathbf{s} \mid \mathbf{R}^T, \mathbf{S}^T)$, can be written as a product of probabilities,

$$\Pr\left(\mathbf{R} = \mathbf{r}, \mathbf{S} = \mathbf{s} \mid \mathbf{R}^T, \mathbf{S}^T\right) = \prod_{i=1}^{n} \Pr\left(R_i \mid R_i^T\right)\Pr\left(S_i \mid S_i^T\right)$$

The individual probabilities $\Pr(R_i \mid R_i^T)$ and $\Pr(S_i \mid S_i^T)$ can be computed using Eq.s (1) and (2). The true values $\mathbf{R}^T$, $\mathbf{S}^T$ are unobserved latent variables which must be integrated out of the joint probability of $\mathbf{R}$, $\mathbf{S}$, $\mathbf{R}^T$, $\mathbf{S}^T$,

$$\begin{aligned}
\Pr(\mathbf{R} = \mathbf{r}, \mathbf{S} = \mathbf{s}) &= \iint \Pr\left(\mathbf{R} = \mathbf{r}, \mathbf{S} = \mathbf{s}, \mathbf{R}^T = \mathbf{u}, \mathbf{S}^T = \mathbf{v}\right)\partial\mathbf{u}\partial\mathbf{v} \\
&= \iint \Pr\left(\mathbf{R} = \mathbf{r}, \mathbf{S} = \mathbf{s} \mid \mathbf{R}^T = \mathbf{u}, \mathbf{S}^T = \mathbf{v}\right)\Pr\left(\mathbf{R}^T = \mathbf{u}, \mathbf{S}^T = \mathbf{v}\right)\partial\mathbf{u}\partial\mathbf{v} \qquad (3) \\
&= \iint \left\{\prod_{i=1}^{n} \Pr\left(R_i \mid R_i^T = u_i\right)\Pr\left(S_i \mid S_i^T = v_i\right)\right\} \Pr\left(\mathbf{R}^T = \mathbf{u} \mid \mathbf{S}^T = \mathbf{v}\right)\Pr\left(\mathbf{S}^T = \mathbf{v}\right)\partial\mathbf{u}\partial\mathbf{v}
\end{aligned}$$

The stock–recruitment model is used for $\Pr(\mathbf{R}^T \mid \mathbf{S}^T)$. We assume,

$$R^T \mid S^T = \mu\left(S^T\right)\exp(\varepsilon_{RPE} - \sigma_{RPE}^2 / 2),\ \ \varepsilon_{RPE} \sim N(0, \sigma_{RPE}^2) \tag{4}$$

We initially assume that the ε_{RPE}'s are independent over time; hence, using Eq. (4),

$$\Pr\left(\mathbf{R}^T = \mathbf{u} \mid \mathbf{S}^T = \mathbf{v}\right) = \prod_{i=1}^{n} \Pr\left(R_i^T = u_i \mid S_i^T = v_i\right) \tag{5}$$

The independence assumption in Eq. (4) will often not be appropriate and we address this issue later. Eq. (5) can be used for $\Pr(\mathbf{R}^T = \mathbf{u} \mid \mathbf{S}^T = \mathbf{v})$ in Eq. (3).

The last component to model is Eq. (3) is Pr($\mathbf{S}^T$). The S^T's will usually be highly dependent over time, but the relationship between S_y^T and S_{y-1}^T can be complex and depends on total mortality rates, growth rates, and maturation rates. We have investigated two approaches. The first was to assume a simple random walk, and the second was to use information from stock assessments. The random walk approach ignored information on the impacts of catches on *S*, and the impacts of changes in growth rates and maturation rates. We also investigated a delayed-difference approach for modelling the relationship between S_y^T and S_{y-1}^T, with *S* "growth rate" information derived from the stock assessment. However, the *S* growth rate information was not independent from the stock assessment estimates of *S*, and we abandoned this approach.

The more statistically rigorous approach is to estimate the SR model within the stock assessment model, and account for ME in S directly in the assessment model. This is the approach we pursue.

Standard errors were derived for SR parameters and MSY RP's using the delta method. Confidence intervals (CI's) were derived by assuming log parameter estimates were normally distributed, and then exponentiating the log CI's.

Internal SR estimation methods

In this approach the SR model is estimated as part of an analytic stock assessment model. The specific assessment model we use is a survey only model, but our conclusions should not be sensitive to the choice of assessment model. The basis of the survey assessment model has been described in Section 9.1 of ICES (2009). We have extended the model in a couple of aspects, particularly because we wish to use the approach to estimate MSY RP's.

SURBA+ background

The survey-only assessment model (SURBA+) provides estimates of trends in stock size and direct estimates of mortality rates based on a time-series of age-based survey indices of stock size ($I_{a,y}$, a=1,…,A, y=1,…Y) and assumptions about survey "catchability" and natural mortality. Population size is modelled using the standard cohort model, $N_{a+1,y+1} = N_{a,y}\exp(-Z_{a,y})$, where $N_{a,y}$ is the beginning of year population size at age a in year y, and $Z_{a,y}$ is the annual total mortality rate. Parameters are estimated using survey indices that are assumed to be related to population size via the observation equation

$$I_{a,y} = q_a N_{a,y} \exp(-p_y Z_{a,y} + \varepsilon_{a,y}) \quad (6)$$

where $p_y Z_{a,y}$ is the fraction of total mortality that occurs before the survey takes place, q's are parameters for survey catchability, and ε's are observation error terms. In Eq. (6), $N_{a,y}$ is projected forward to the time of the survey by applying the fraction of total mortality.

The total mortality rate is split into a user supplied contribution due to natural mortality ($M_{a,y}$), and an estimated contribution due to fishing ($F_{a,y}$). The basis of SURBA+ is a simple separable model for fishing mortality, $F_{a,y} = s_a f_y$, where s_a is the fishery selectivity for different ages, which is assumed to be year-invariant. The F year effects, f_y, are identified by constraining $s_a = 1$ for some age a_{full} that one expects is the first age fully recruited to the fishery. We also assume that selection on the oldest two ages is equal.

There is confounding between q_a's and s_a's in a SURBA model (e.g. Section 4.1.2.2 in ICES 2008b). To remove this confounding, values for q's are usually supplied by the user (i.e. assumed or derived from external sources). Hence, SURBA provides population size estimates that are relative to the assumed scale of the survey q's. SURBA is a highly parameterized model, even when q values are fixed, and it is useful to control the variation in some parameter values.

We use a "random walk" to control or smooth one of the between-year variation in f_y 's. This is sensible if one expects that the true fishing mortality rates do not vary substantially from year to year, which makes sense for stocks like the ones in our examples, in which the fishery is not based mostly on recruitment and quota's do not change much from year to year. The random walk is

$$\log(f_y) = \log(f_{y-1}) + \delta_y^{(F)}, \quad (7)$$

where $\delta_y^{(F)}$ are independent $N(0, \sigma_F^2)$ random error terms. The variance σ_F^2 could be user specified but a more objective modelling approach is to estimate σ_F^2 and let the data decide how much smoothing is appropriate. This is easy to do using the ADMB random effects module (ADMB-RE), which uses the marginal likelihood, in which the δ random effects are numerically "integrated out", for inference about fixed effect parameters like σ_F^2. A fixed-effect (i.e. not random) mean parameter may be specified to start the random walk, but this is not necessary in ADMB-RE.

As a result of moratoria on directed fishing, fishing mortality was very low during 1994–1996 for both of the stocks in our examples. To account for this, we did not include the f_y 's during 1994–1996 as part of the random walk. We simply set these values to be 0.05. We re-started the random walk in 1997. Hence, the specific random walk model we used in our examples is

$$\log(f_y) = \left\{ \begin{array}{ll} \log(0.05), & y = 1994,\ 1995,\ 1996, \\ \log(f_{y-1}) + \delta_y, & \text{otherwise.} \end{array} \right\}$$

Another problem we had to contend with in our examples was survey year effects. The surveys in that area can be highly variable in which there can be large increases or decreases in survey catch rates for all ages. The magnitude of the changes can be well beyond what is possible in the stock. Not accounting the year effects resulted in high estimates of σ_F^2, unrealistic predictions of F in some years, and residual year effects (i.e. correlated errors). This problem of survey year effects is fairly common in stock assessments. We addressed this problem by modifying the observation equation (i.e. Eq. 6),

$$I_{a,y} = Q_{a,y} N_{a,y} \exp(-pZ_{a,y} + \varepsilon_{a,y}), \tag{8}$$

where $\log(Q_{a,y}) = \log(q_a) + \tau_y$ and τ_y are independent $N(0, \sigma_Q^2)$ random year effects. Unfortunately there is some confounding between the magnitude of σ_Q^2 and σ_F^2 so we simply set $\sigma_Q = 0.25$ which seemed to result in plausible estimates of F and little to no correlation in residuals.

The other model component to specify is the selectivity, s_a. We expect that s_a varies smoothly as a function of age. A variety of parametric models have been used for this purpose; however, sensitivity to parametric assumptions is always a concern. We decided to also use a random effects approach to produce smooth nonparametric estimates of s_a. At younger ages we expect $\log(s_a)$ to increase roughly linearly with age, but at older ages we expect much less change in s_a's. To accommodate this type of variation we used the following random effects model:

$$\log(s_a) = \begin{Bmatrix} 0.5\left[\log(s_{a+1}) + \log(s_{a-1})\right] + \xi_a, & a \le a_{full}, \\ \log(s_{a-1}) + \xi_a, & \text{otherwise}, \end{Bmatrix} \tag{9}$$

where ξ_a is a normal distribution error term. This model penalizes against first-order differences at older ages, $\{\log(s_{a+1}) - \log(s_a)\}^2$, and will favour constant selectivity unless there is "strong" evidence in the data for a trend. The strength of the evidence is determined via the improvement in fit to the data vs. the likelihood-cost of having $\mathrm{Var}(\xi_a) = \sigma_S^2 \ge 0$ in the likelihood. At younger ages the penalty is $\{\log(s_{a+1}) + \log(s_{a-1}) - 2\log(s_a)\}^2$ which favours log-linear selectivities; that is, the penalty is zero when s_a is log-linear in a. In some preliminary analyses the estimates of s_a 's seemed too variable (i.e. σ_S^2 too large) so we constrained $\sigma_S = 0.2$.

Recruitment in SURBA+

Similar to Eq. (1), we model recruitment as a stochastic function of parental stock size,

$$R^T \mid S^T = \mu\left(S^T\right)\exp(\varepsilon_{RPE}),\ \ \varepsilon_{RPE} \sim N(0, \sigma_{RPE}^2) \tag{10}$$

Note that we have not bias-corrected the process error in $\exp(\varepsilon_{RPE})$. Parental stock size (S^T) is estimated from the assessment model SSB (S; i.e. sum of mid-year biomass-at-age times maturity estimates) and other measurement error,

$$S^T = S\exp(\varepsilon_{SME}),\ \ \varepsilon_{SME} \sim N(0,\sigma^2_{SME}) \tag{11}$$

We can estimate σ^2_{RPE} because we assume that the survey observation error variances in Eq.s (6) and (8) are the same for all ages. Unfortunately σ^2_{SME} is confounded with σ^2_{RPE} and we cannot reliably estimate these two variance parameters separately. Hence, we cannot estimate the size of the S ME's (ε_{SME}) and the R PE's (ε_{RPE}) separately. Virtually the same stock assessment model fits may be obtaining using a wide range of values for σ^2_{SME}, and some choices can have a large impact of RP's. The best we can do is estimate the RP's over a range of values for σ^2_{SME}. Additional information is required to estimate RP's more precisely. Empirical Bayes estimates of the ε_{RPE}'s and ε_{SME}'s can be used with Eq.s (10) and (11) to obtain more specific estimates of R and S for the assessment model time period.

Note that if $\sigma^2_{SME} = 0$ then the assessment model has no S ME. We refer to the SR model estimated this way as internal. If σ^2_{RPE} is large then the SR model has little to no affect on the assessment model. The size of each cohort in the assessment model is then estimated independently of parental stock size. Estimates of R and S obtained this way (e.g. XSA, ADAPT) are often used to estimate the SR, which is a two-stage procedure. We refer to such SR model estimates as external. In the following examples we examine differences in BH SR parameters and derived MSY RP's when the SR model is estimated external to the assessment model (EXT), internal to the assessment model (INT), and internal with ME. Errors may be independent and identically distributed (IID) or autocorrelated in some way (AR).

Results

Example 1: Atlantic cod (*Gadus morhua*) in NAFO Subdivisions 3Ps

SURBA has been used recently (e.g. DFO 2010) in assessments of Atlantic cod (*Gadus morhua*) in NAFO Subdivision 3Ps which is located off the south coast of Newfoundland, Canada. The SURBA+ model was applied to the DFO survey index for the years 1983–2009 and ages 1–12. M was assumed to be equal 0.2 for all ages and years. We fixed q_a = 0.154 ,0.462 ,and 0.923 for a = 1,…,3 and q_a = 1 for a>3. These values have been used in the most recent assessment for this stock. Estimates of population size are relative to these assumptions about survey catchability. We assumed the fully recruited age to be fishery was six (i.e. a_{full} = 6).

We illustrate some basic analytic assessment results for the EXT SURBA+ model run. Stock size estimates are shown in Figure 2. The current limit reference point for this stock is based on Brecovery, which is SSB in 1994. Estimates of stock status relative to the LRP are shown in Figure 3. Average F's are shown in Figure 4 and fishery selectivity in Figure 5, rescaled so that the maximum is one. Predictions of survey year effects are shown in Figure 6. These year-effects are the reason why the predicted survey indices (Figure 7) can vary substantially from year to year, similar to the survey indices. Residual patterns (Figure 8) do not indicate serious assessment model lack of fit.

We also fitted a SURBA+ model in which the age pattern in fishing mortality could smoothly change over time. Although the results suggested that selectivity may be more slightly domed since the fishing moratorium ended in 1997 (see ICES, 2009), this model did not explain much additional variability and we have not pursued it further.

The negative loglikelihood (nll) for the EXT Surbap+ model was 286.846. The nll for the INT model was 270.520 which is a substantial reduction and indicates the BH SR model explains a significant amount of variation, albeit with considerable process error ($\hat{\sigma}_{RPE} = 0.29$). The unconstrained estimate of σ_{SME} from the Surba+ ME model was close to zero with nll = 270.520, which is very close to the INT nll as expected because when $\sigma_{SME} = 0$ these models are the same. The nll when σ_{SME} was fixed at 0.25 was only slightly greater, nll = 270.569, and $\sigma_{RPE} = 0.26$ which indicates the confounding between σ_{RPE} and σ_{SME}.

The BH SR model estimates of MSY RP's from the EXT, INT, and ME ($\sigma_{SME} = 0.25$) Surba+ models are shown in Figure 9. The EXT and INT SR models and RP's are similar, although there is less variation from the BH model in the INT predictions of R, as expected. The INT and ME SR models and RP's are also similar. However, when $\sigma_{SME} = 0.4$ (nll = 270.783) the differences in the ME RP's are larger (Figure 10).

The predicted ME errors in SSB are somewhat consistent with concerns about declines in the age at maturity for 3Ps cod. The concerns are that fish that mature at young ages produce fewer eggs of lower quality, and do not contribute as much to SSB as older fish. The trend in the predicted errors from the ME model is negative when proportion mature at young ages increases (Figure 11). Negative errors mean that SSB is reduced, indicating less spawning potential than estimated by the sum of mature biomass at age.

Example 2: American plaice (*Hippoglossoides platessoides*) in NAFO Subdivision 3Ps

The SURBA+ model was applied to the DFO survey index for the years 1983–2009 and ages 5–15. M was assumed to be equal 0.2 for all ages and years. We fixed $q_a = 0.6$ and 0.8 for $a = 5$ and 6, and $q_a = 1$ for $a > 6$. These were the values that were close to one but also accounted for the partial recruitment of these ages to the survey – as evidenced by negative raw survey Z's and Surba+ residual patterns with age. Estimates of population size are relative to these assumptions about survey catchability. We assumed the fully recruited age was seven (i.e. $a_{full} = 7$). Because there are only two ages in the model less than a_{full} in Eq. (9) we used the first-order random walk for all ages.

We illustrate some basic analytic assessment results for the EXT SURBA+ model run. Stock size estimates are shown in Figure 12. The SSB in 1994 was the minimum in the time-series. Estimates of SSB relative to the 1994 value are shown in Figure 13. The purpose of this figure is to describe the trend in SSB relative to the 1994 level. Average F's are shown in Figure 14 and fishery selectivity in Figure 15, rescaled so that the maximum is one. Predictions of survey year effects are shown in Figure 16. These year-effects are the reason why the predicted survey indices (Figure 17) can vary substantially from year to year, similar to the survey indices. Residual patterns (Figure 18) do not indicate serious assessment model lack of fit.

The negative loglikelihood (nll) for the EXT Surbap+ model was 227.483. The nll for the INT model was 212.617 which is a substantial reduction and indicates the BH SR model explains a significant amount of variation, albeit with considerable process error ($\hat{\sigma}_{RPE} = 0.31$). The unconstrained estimate of σ_{SME} from the Surba+ ME model was large ($\hat{\sigma}_{SME} = 0.69$) with nll = 211.920, which is close to the INT nll where $\sigma_{SME} = 0$. This demonstrates the confounding between σ_{RPE} and σ_{SME}. The nll when σ_{SME} was fixed at 0.25 was only slightly greater, nll = 212.561.

The BH SR model estimates models and estimates of MSY RP's from the EXT, INT, and ME ($\sigma_{SME} = 0.25$) Surba+ are shown in Figure 19. The RP's from the three models are similar. However, when $\sigma_{SME} = 0.69$ the differences in the ME RP's are much larger (Figure 20).

Discussion

There may be additional sources of error in determining parental stock size, apart from errors in the components of the common SSB calculation. We refer to these errors as parental stock size measurement errors, or SME. These additional sources of error may be due to, for example, trends in sex ratio and fecundity, or possible effects from changes in reproductive potential related to changes in age composition. The variance of the SME's cannot be estimated in a stock assessment model unless additional and specific information is available about the size of these effects. This information will usually not be available, and the best we can do is provide a sensitivity analysis of the impact of SME's pm assessment model results. The main impact in our examples was on the stock–recruit parameter estimates and MSY reference points. If the SME's are large then reference points can be substantially different from those obtained assuming no SME's. In the 3Ps cod example, Bmsy decreased when SME was large, and Fmsy increased slightly. In the 3Ps American plaice example, Bmsy increased substantially when SME was large, and Fmsy decreased. It would be useful to provide profile plots of how estimated MSY reference points are affected by assumptions about SME.

Further extensions of the methodology are still required. As mentioned earlier, it seems likely that SME's will be autocorrelated over time. It may also not be possible to estimate the amount of autocorrelation, in which case a sensitivity analysis into this aspect of SME is also necessary.

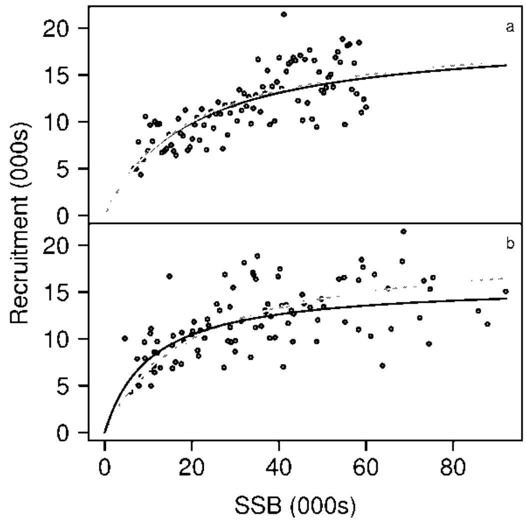


Figure 1. Simulation example, with 100 stock (SSB) and recruitment observations generated from a Beverton–Holt model (grey line in both panels) with a slope at the origin of 1.0 and maximum recruitment of 20 000. Estimated models are shown as black lines. Panel a): Lognormal recruitment observations were generated with a coefficient of variation of 0.2. Panel b): The same recruitment observations as in panel a), but additional lognormal measurement errors with a coefficient of variation of 0.4 were added to the SSB's.

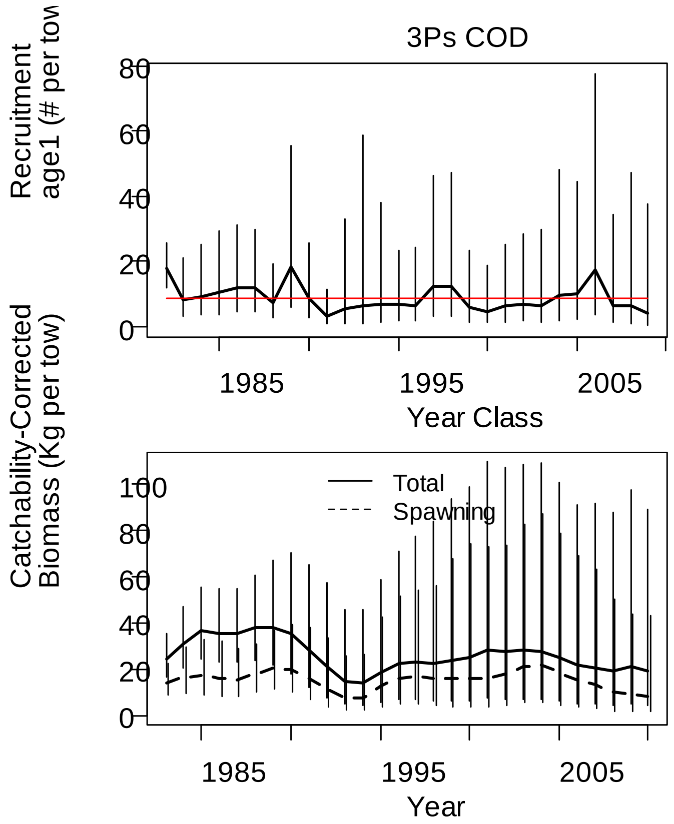


Figure 2. Estimates of recruitment (top left panel) and biomass (bottom left panel). The red line indicates the time-series mean. Vertical lines indicate 95% confidence intervals.

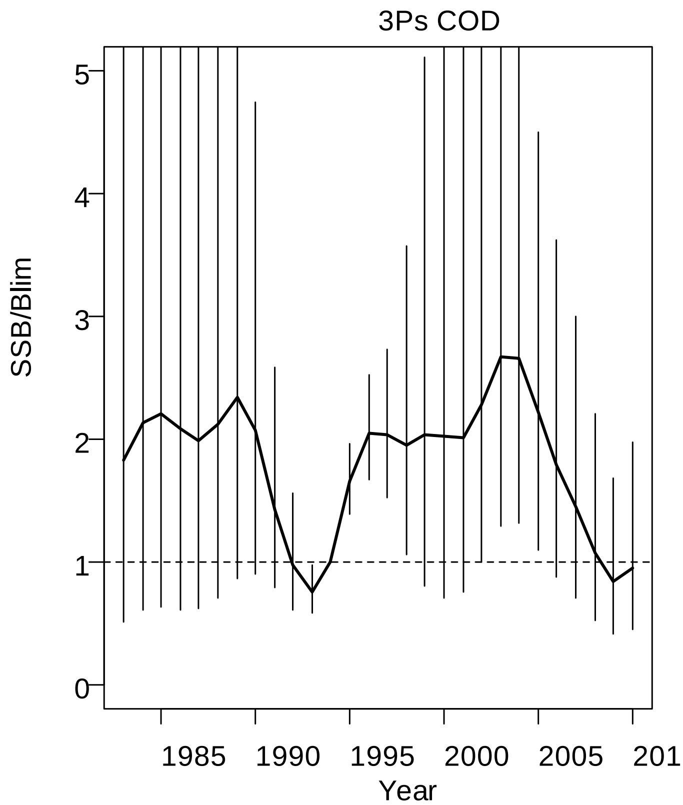


Figure 3. Estimates of SSB relative to 1994 values, which for SSB is the limit reference point for this stock. Vertical lines indicate 95% confidence intervals. Dashed references lines at one are shown.

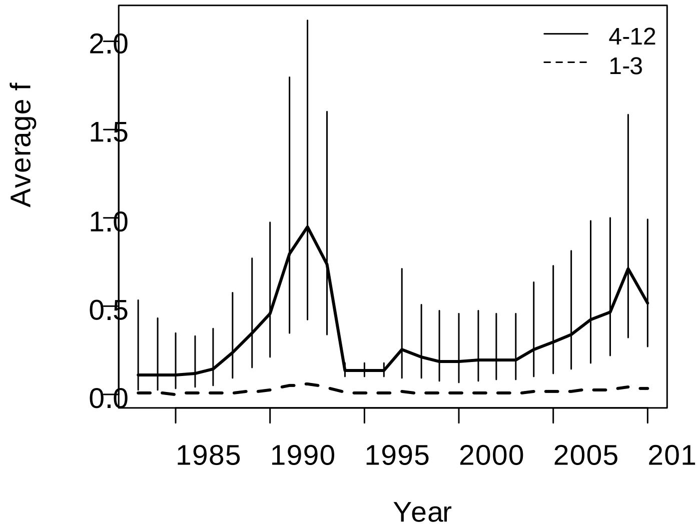


Figure 4. Average fishing mortality for ages 4–12 (solid line) and 1–3 (dashed line). Vertical lines indicate 95% confidence intervals.

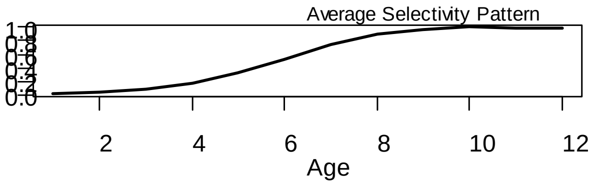


Figure 5. Predicted fishery selectivity.

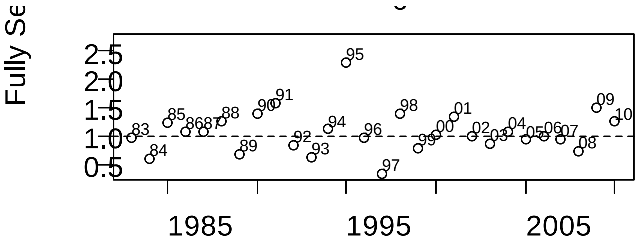


Figure 6. Year effects in survey catchability (Q).

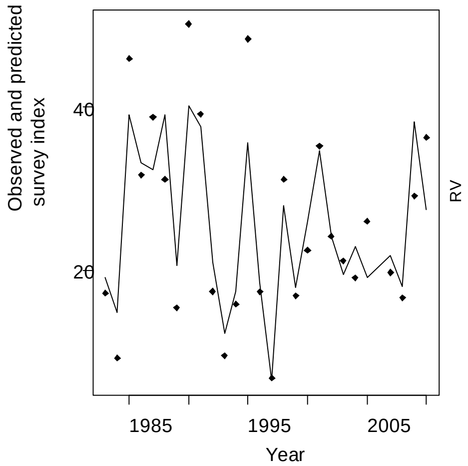


Figure 7. Sum of observed (points) and predicted (lines) survey indices each year. Predicted indices have been corrected for the log transformation bias.

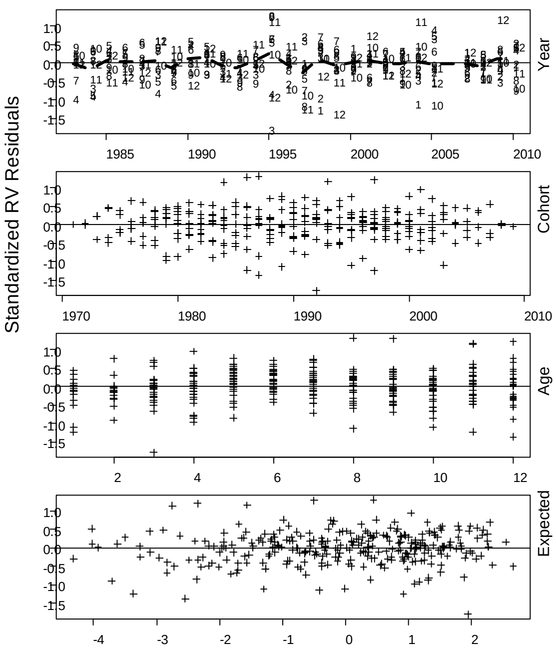


Figure 8. Standard residuals vs. year, age, cohort, and predicted value. The dashed line in the top right panel indicates the average residual each year.

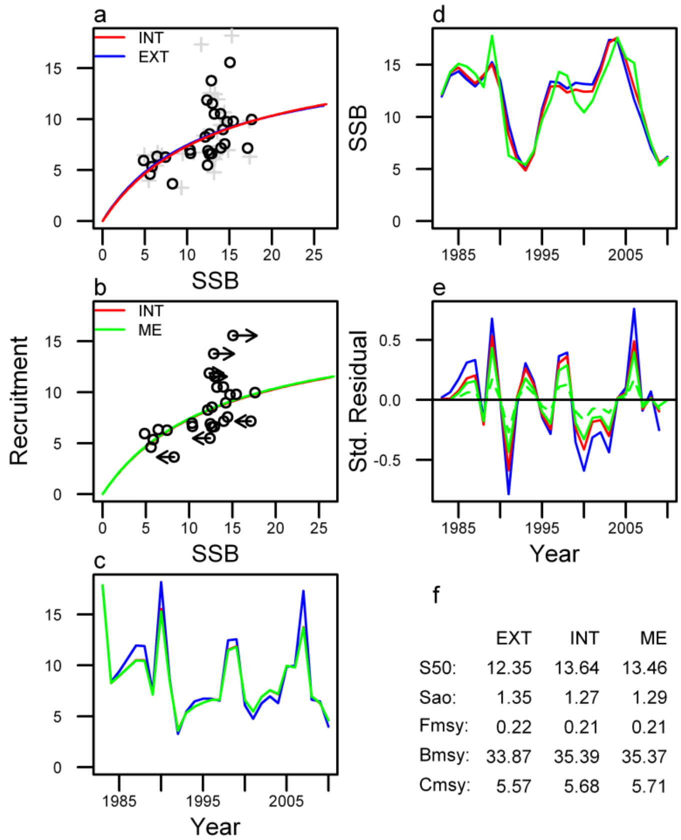
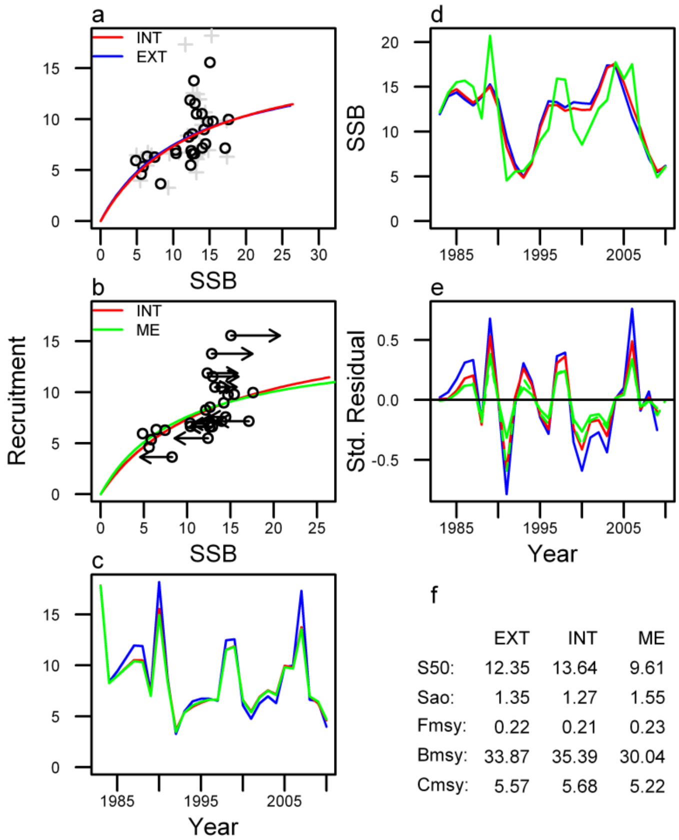


Figure.9. Beverton–Holt stock–recruit (SR) model fits to 3Ps cod (*Gadus morhua*) data. Panel a) External (EXT; blue line and grey +'s) and internal (INT; red line and o's) SR fits and predictions of SSB and recruitment (R). The INT model has no measurement error (ME) in SSB. Panel b) INT with ME (sd=0.25; green line) and without ME (red lines and o's). Arrows indicate the predicted SSB's from the ME model. ME standardized residuals. Panel c) R time-series from the EXT, INT, and ME models. Panel d) SSB time-series from the EXT, INT, and ME model. Panel e) Residual time-series. The solid lines correspond to the models in panels a) and b). The dashed line indicates the ME SSB residuals. Panel f) SR parameters and MSY reference points.



	EXT	INT	ME
S50:	12.35	13.64	9.61
Sao:	1.35	1.27	1.55
Fmsy:	0.22	0.21	0.23
Bmsy:	33.87	35.39	30.04
Cmsy:	5.57	5.68	5.22

Figure 10 ME sd=0.4. See Figure 9 for other details.

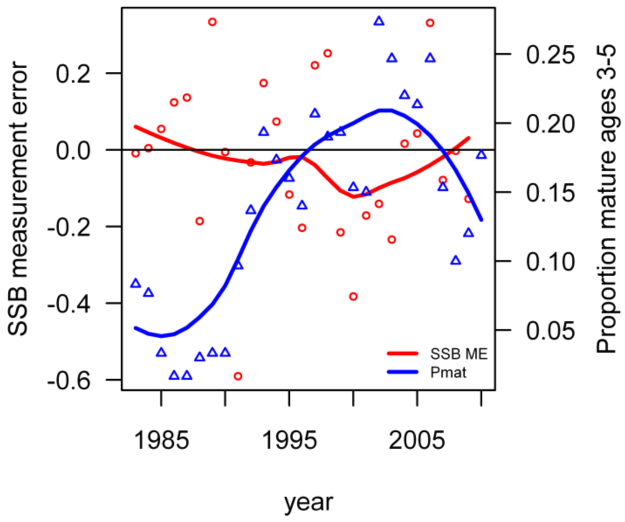


Figure 11. Time-series of average proportion mature at ages 3–5 (blue triangles), and the predicted errors in SSB (red circles). The solid lines are loess smoothed trends in the estimates.

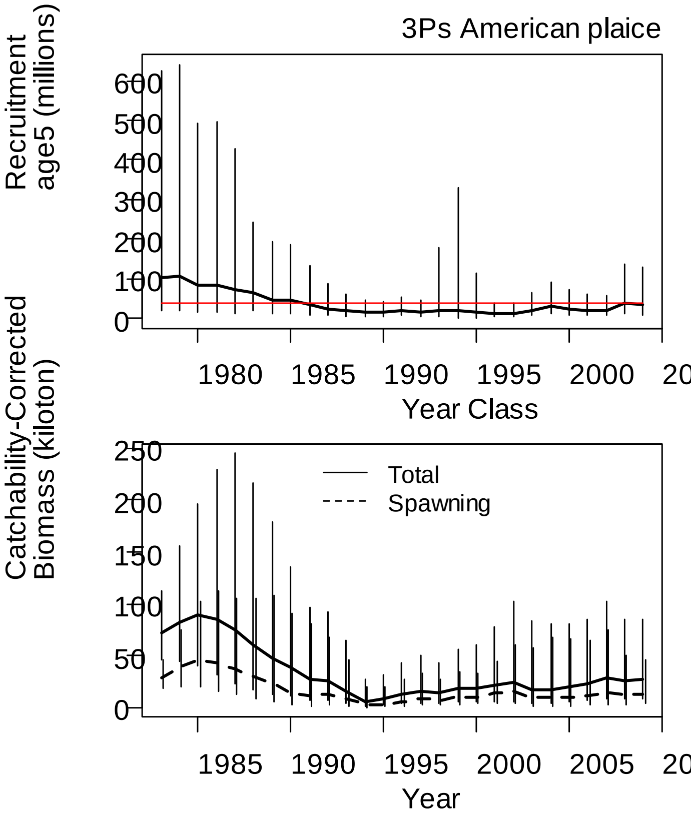


Figure 12. Estimates of recruitment (top left panel) and biomass (bottom left panel). The red line indicates the time-series mean. Vertical lines indicate 95% confidence intervals.

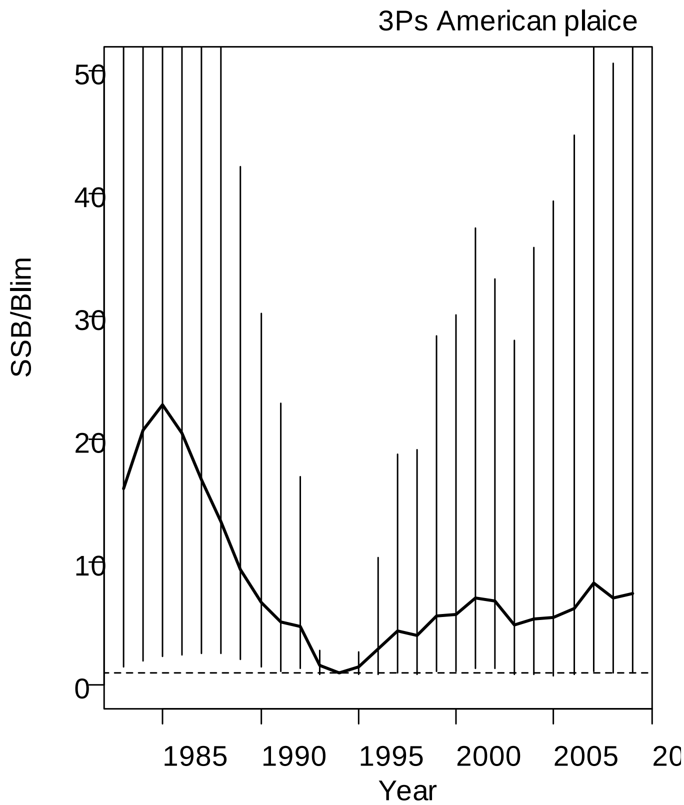


Figure 13. Estimates of SSB relative to 1994 values. This is not the limit reference point for this stock – the figure label is wrong. Vertical lines indicate 95% confidence intervals. Dashed references lines at one are shown.

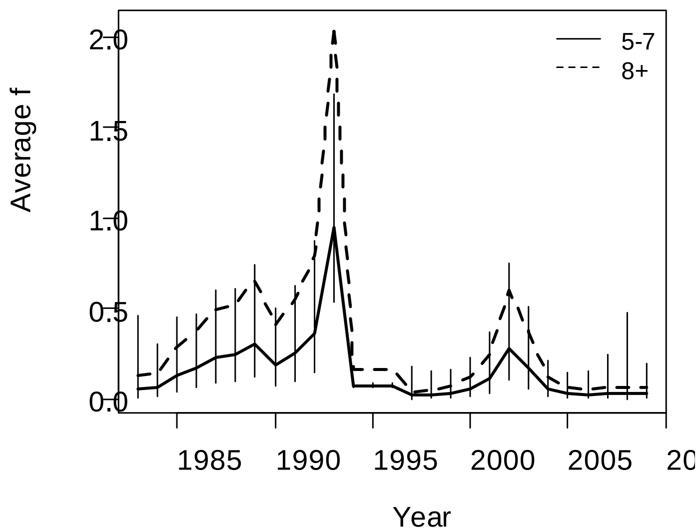


Figure 14. Average fishing mortality. Vertical lines indicate 95% confidence intervals.

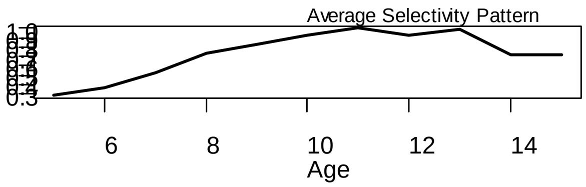


Figure 15. Predicted fishery selectivity.

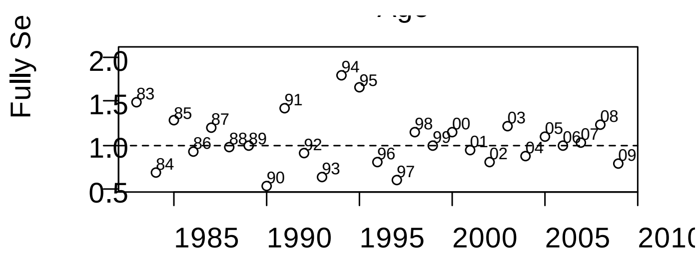


Figure 16. Year effects in survey catchability (Q).

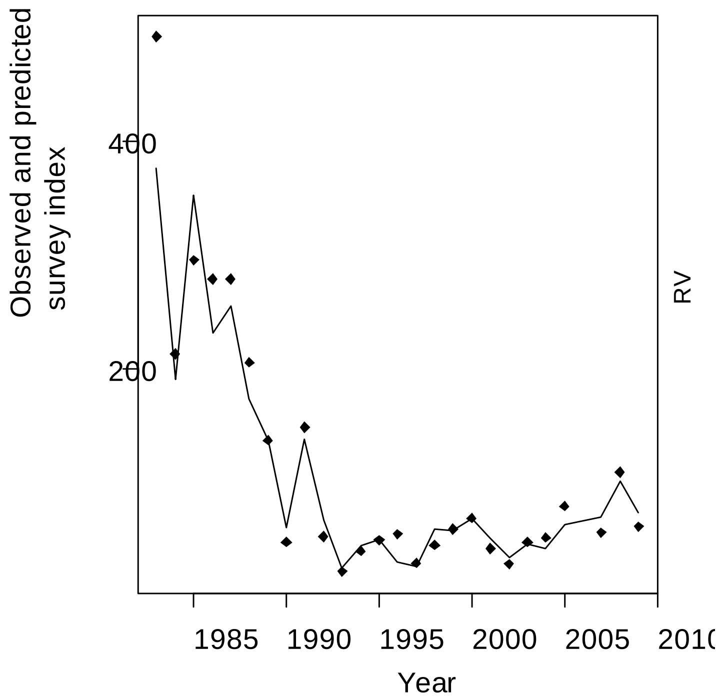


Figure 17. Sum of observed (points) and predicted (lines) survey indices each year. Predicted indices have been corrected for the log transformation bias.

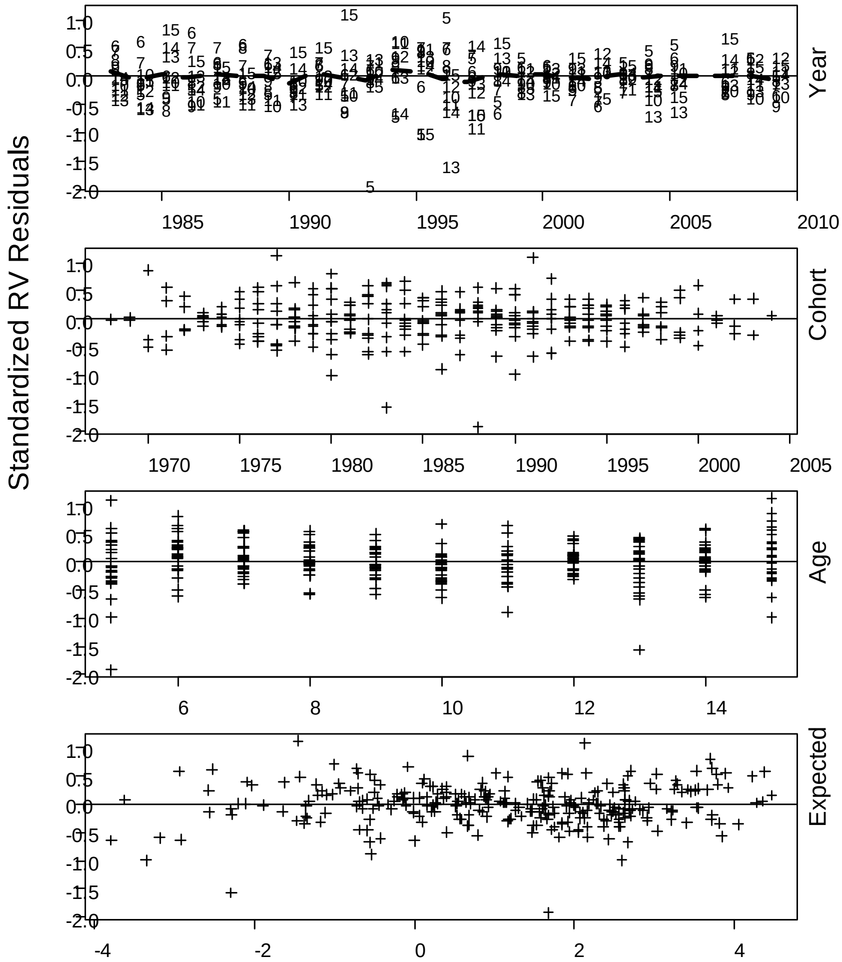


Figure 18. Standard residuals vs. year, age, cohort, and predicted value. The dashed line in the top right panel indicates the average residual each year.

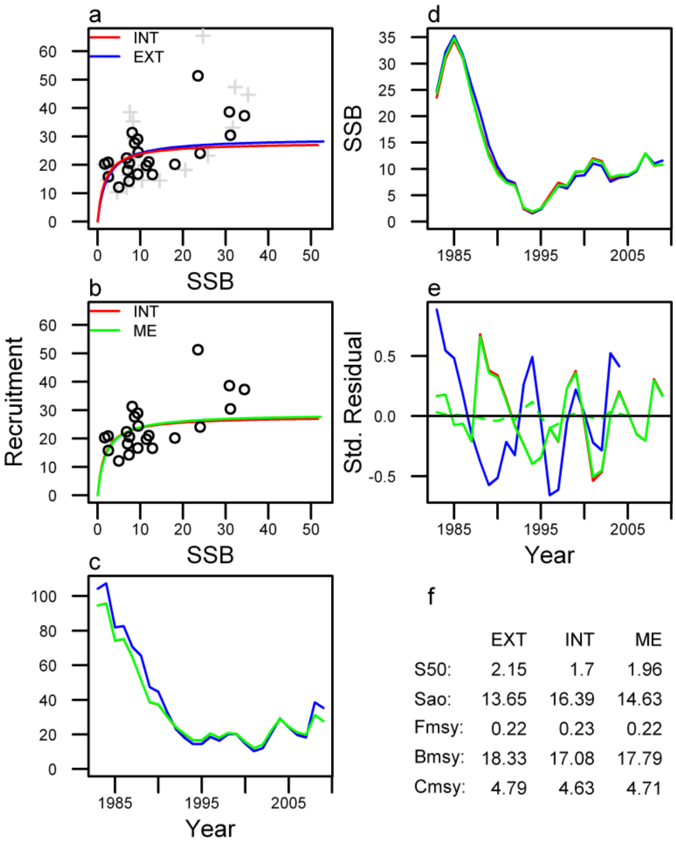


Figure 19. Beverton–Holt stock–recruit (SR) model fits to 3Ps American plaice (*Hippoglossoides platessoides*) data. Panel a) External (EXT; blue line and grey +'s) and internal (INT; red line and o's) SR fits and predictions of SSB and recruitment (R). The INT model has no measurement error (ME) in SSB. Panel b) INT with ME (sd=0.25; green line) and without ME (red lines and o's). Arrows indicate the predicted SSB's from the ME model. ME standardized residuals. Panel c) R time-series from the EXT, INT, and ME models. Panel d) SSB time-series from the EXT, INT, and ME model. Panel e) Residual time-series. The solid lines correspond to the models in panels a) and b). The dashed line indicates the ME SSB residuals. Panel f) SR parameters and MSY reference points.

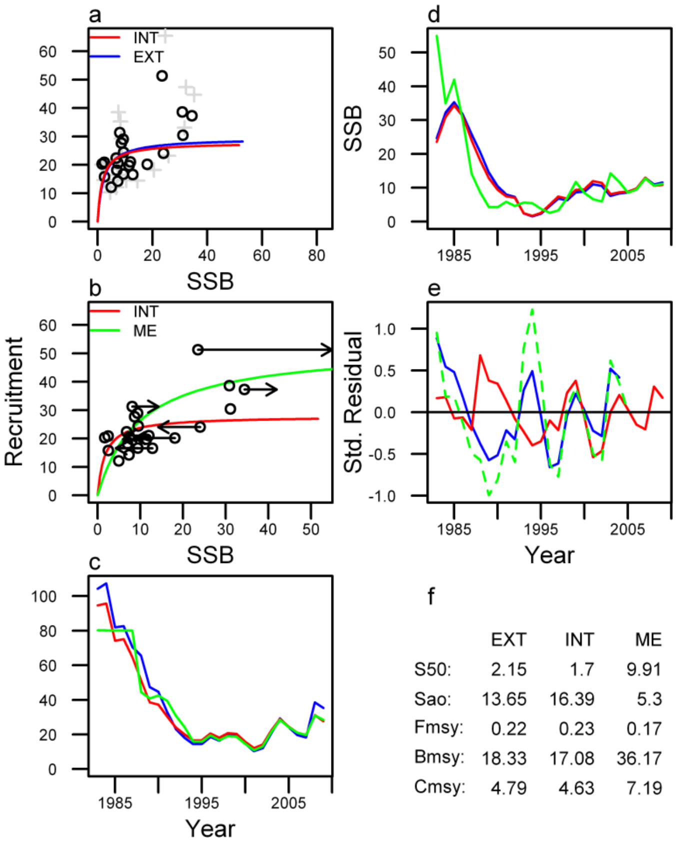


Figure 20. ME sd=0.63; the estimated value See Figure 19 for other details.

References

ADMB Project. 2009 AD Model Builder: automatic differentiation model builder. Developed by David Fournier and freely available from admb-project.org.

Beverton, R. J. H., and Holt, S. J. 1957. On the dynamics of exploited fish populations. Chapman and Hall, London.

Cadigan, N. 2009. Sensitivity of common estimators of management parameters derived from stock–recruit relationships. Fisheries Research, 96: 195–205.

Carroll, R. J., Ruppert, D., Stefanski, L. A., Crainiceanu, C. M. 2006. Measurement Error in Nonlinear Models. A Modern Perspective, Second Edition. Chapman and Hall/CRC.

Conn, P. B., Williams, E. H., and Shertzer, K. W. 2010. When can we reliably estimate the productivity of fish stocks? Canadian Journal of Fisheries and Aquatic Sciences, 67: 511–523.

DFO. 2010. Stock Assessment of Subdivision 3Ps cod, October 2010. DFO Can. Sci. Advis. Sec. Sci. Advis. Rep. 2010/067.

Hilborn, R. 2010. Pretty Good Yield and exploited fishes. Marine Policy, 34: 193–196.

Hilborn, R., and Stokes, K. 2010. Defining Overfished Stocks: Have We Lost The Plot? Fisheries, 35: 113–120.

ICES. 2009. Report of the Working Group on Methods of Fish Stock Assessment (WGMG). ICES CM 2009/RMC:12.

Jiao, Y., Schneider, D., Chen, Y., and Wroblewski, J. 2004. An analysis of error structure in modeling the stock–recruitment data of gadoid stocks using generalized linear models. Canadian Journal of Fisheries and Aquatic Sciences, 61: 134–146.

Kehler, D. G., Myers, R. A., and Field, C. A. 2002. Measurement error and bias in the maximum reproductive rate for the Ricker model. Canadian Journal of Fisheries and Aquatic Sciences, 59: 854–864.

Kope, R. 2006. Cumulative effects of multiple sources of bias in estimating spawner-recruit parameters with application to harvested stocks of Chinook salmon (*Oncorhynchus tshawytscha*). Fisheries Research, 82: 101–110.

Legović, T., Klanjčšek, J., and Geček, S. 2010. Maximum sustainable yield and species extinction in ecosystems. Ecological Modelling, 221: 1569–1574.

Michielsens, C. G. J., and McAllister, M. K. 2004. A Bayesian hierarchical analysis of stock–recruitment data: quantifying structural and parameter uncertainties. Canadian Journal of Fisheries and Aquatic Sciences, 61: 1032–1047.

Morgan, M. J., A. Perez-Rodriguez, F. Saborido-Rey. 2011. Does increased information about reproductive potential result in better prediction of recruitment? Can. J. Fish. Aquat. Sci., 68:1361–1368.

Needle, C. L. 2002. Recruitment models: diagnosis and prognosis. Reviews in Fish Biology and Fisheries, 11: 95–111.

Quinn, T. J., Deriso, R. B., 1999. Quantitative fish dynamics. Oxford University Press, New York.

Ricker, W. E. 1954. Stock and recruitment. Journal of the Fisheries Research Board of Canada, 11: 559–623.

Sissenwine, M. P., and Shepherd, J. G. 1987. An Alternative Perspective on Recruitment Overfishing and Biological Reference Points. Canadian Journal of Fisheries and Aquatic Sciences, 44: 913–918.

Walters, C. J. 1990. A parital bias correction factor for stock–recruitment parameter estimation in the presence of auto-correlated environmental effects. Canadian Journal of Fisheries and Aquatic Sciences, 47: 516–519.

Walters, C. J., and Ludwig, D. 1981. Effects of measurement errors on the assessment of stock--recruitment relationships. Canadian Journal of Fisheries and Aquatic Sciences, 38: 704–710.

Annex 6: Estimating σ_R

In the context of discussing WD 6 (Annex 5), it was noted that the estimation of the recruitment variation parameter σ_R in a Maximum Likelihood Estimation (MLE) setting when a stock–recruit model forms part of the assessment model, and the recruitment residuals ε_y appear in a penalised likelihood term

$$-\ln L = n\ln(\sqrt{2\pi}\sigma_R) + \frac{1}{2\sigma_R^2}\sum_y \varepsilon_y^2$$

leads to a maximum likelihood estimate of $\sigma_R = 0$ if no additional constraints are placed on the model. This did not appear to happen in the state-space model presented in WD 6, and a discussion arose as to why this was the case for that model, which treats process errors as random effects that are integrated out.

Part of the problem stems from the fact that the penalised likelihood given in the above equation is not a "true" likelihood, in the sense that ε_y measures the distance between what are essentially two model estimates (the parametric stock–recruit curve and the estimated recruitment in year y), instead of measuring the distance between (fixed) data values and model estimates. The reason why no sensible MLE exists is that the above equation goes to $-\infty$ when ε_y exactly equals 0 and σ_R is positive but tends towards zero. In this case, it is possible that all the ε_y equal zero exactly, given that these are treated as unknown parameters to be estimated. If, on the other hand, annual recruitments were observed (*i.e.* were treated as data, rather than as parameters to be estimated), then the ε_y could not all be exactly equal to zero (because it would be extremely unlikely that the stock-recruitment relationship would be able to match all the observed annual recruitments exactly). Therefore, ε_y^2 would be strictly positive, at least for some years, in which case the above equation for $-\ln L$ would be minimised for some finite (and strictly positive) value of σ_R.

When the annual recruitment as reflected by ε_y is treated as unknown (rather than as known data), the equation above is not really a likelihood function for recruitment but, instead, a probability distribution for it (or, using a different terminology, a random effect). In that case, the equation is implicitly saying that the annual recruitment is a random variable (varying around the stock-recruitment relationship and with the magnitude of the variability controlled by σ_R). Hence, annual recruitments cannot be treated just as parameters to be estimated by ML, since the equation above is not a likelihood. Instead, they must be integrated out of the model, using their probability distribution, and apart from the stock recruitment function parameters themselves, the only other recruitment-related parameter left in the model (which is, indeed, a real parameter) is σ_R. Because annual recruitments have been integrated out of the model using their entire probability distribution (and not just estimating $\varepsilon_y = 0$), the ML estimate of σ_R no longer suffers from the problem caused by ε_y being exactly equal to zero and the ML estimate of σ_R will typically be a finite (and strictly positive) value. In this case, a kind of "posterior" probability distribution can be subsequently obtained for the annual recruitment values, which will be different from the original one input in the model (*i.e.* the one at the top of this section), as the "posterior" distribution of recruitment will be centred at different values on different years depending on the information coming from other parts of the model and from the various data sources used to fit the model.

The above provides an explanation as to why the $\sigma_R = 0$ problem does not arise for the state-space model presented in WD 6, as the implementation there uses the Random Effects module of ADMB to integrate out the random effects. Note that the pole (infinite value) at σ_R = 0 in the integrand provided by the equation above will not cause the likelihood profile in σ_R obtained after integrating out the recruitments to increase indefinitely as σ_R tends to zero; this is because the very large value of the integrand in this region is more than compensated by the rate at which the volume of the integration phase space over the recruitments approaches zero in this limit, so that a non-zero MLE of σ_R will result. The bottom line here is that, in order to avoid pathological behaviour, random effects should be treated specifically as such, and not merely as parameters to be estimated via ML. The practical problem is that integrating out random effects can be difficult: there is often no analytical solution to the integral and numerical integration must be used. Actually implementing this numerical integration in ADMB is not straightforward when models become more complicated (i.e. less well approximated by linear and Normal assumptions).

Considering this problem during a recent stock assessment workshop, Smith *et al.* (2011) recommended that an assessment model should be treated as a random effects model and the process errors integrated out. However, because analytical integration could be very complicated in many cases, an alternative approach is to include a prior distribution on σ_R in the estimation to keep the "maximum likelihood" estimate of σ_R away from zero and with a positive definite Hessian, and then to drop this prior when implementing an MCMC algorithm (that Hessian then having provided a sensible jump function for this algorithm).

Reference

Smith, A. D. M., Fernandez, C., Parma, A., and A. E. Punt. 2011. International Review Panel report for the 2011 International Fisheries Stock Assessment Workshop. 28 November-2 December 2011, UCT: 19pp. [accessed online, 13 December 2011: http://www.mth.uct.ac.za/maram/workshops.php].