Fisheries Research 108 (2011) 153–162

Contents lists available at ScienceDirect

Fisheries Research

journal homepage: www.elsevier.com/locate/fishres

Comparison of stock management with production, difference, and age-structured models using operating models

Jan Horbowy*

Sea Fisheries Institute, Dpt of Fisheries Resources, Kołłątaja 1, 81-332 Gdynia, Poland

ARTICLE INFO

Article history:
Received 7 May 2010
Received in revised form
13 December 2010
Accepted 13 December 2010

Keywords:
Stock
Assessment
Model
Management
MSE

ABSTRACT

Management strategy evaluation (MSE) was used to test the assessment and management performance of three assessment methods in combination with harvest control rules. The assessment procedures considered were: the eXtended Survivors Analysis (XSA), the Schaefer production model, and the difference model. Four HCRs were considered: first, fishing mortality was set on the basis of the relationship between the current biomass and a reference biomass; second, fishing mortality was gradually reduced (or increased) until it reached a required target; and the third and the fourth HCRs were similar to the first and second but with imposed TAC constraints. The stock that was generated in the operating model (OM) resembled the eastern Baltic cod stock. For the XSA assessment, two options were used: XSA with default shrinkage of terminal fishing mortality to the average of the estimates, and XSA with low shrinkage. The simulations showed that for stock assessment, the XSA models performed much better than the difference and Schaefer models. However, for the data tested, the difference and Schaefer models performed somewhat better in terms of management performance than the XSA models, especially the XSA model with default shrinkage.

1. Introduction

Fishery management usually requires estimates of historical stock sizes and predictions of future catches and stock size under various fishing mortality or fishing effort conditions. To make these estimations and predictions, scientific advisory bodies apply mathematical models that are typically age-structured and have different levels of complexity. ICES, the International Council for the Exploration of the Sea, is the main advisory body on fishery management for stocks exploited in most European waters. The standard ICES procedure for estimating biomass and fishing mortality for stocks in the northeast Atlantic involves the use of the eXtended Survivors Analysis (XSA; Shepherd, 1999) or Integrated Catch Analysis (ICA; Patterson, 1998). Surplus production models (e.g., Schaefer, 1954) are used less often for fish stock management but have been used within the ICES community for Greenland halibut and anglerfish (ICES, 2009), and have been used routinely by ICCAT (International Commission for the Conservation of Atlantic Tunas) for tuna and tuna-like species (ICCAT, 2010). Difference models were developed as an alternative to production models (Deriso, 1980; Horbowy, 1992) and may be useful for stocks in which recruitment undergoes larger variations because recruitment is modelled separately or taken as an external variable.

Age-structured models require historical age data from commercial catches and surveys, which are costly to obtain. In addition, for some species, there are problems with age determination (e.g., cod in the Baltic, ICES (2006a)). Thus, it would be interesting to compare the results of a stock management method that employs age-structured models with one that is based on simpler production or difference models that demand less data.

Recently, computer-intensive methods have been increasingly used in stock management studies, and a management strategy evaluations (MSEs) approach has been applied (e.g., Patterson et al., 2001; Kell et al., 1999, 2005; Dichmont et al., 2006; Rademeyer et al., 2007). The MSE typically consists of an operating model (OM), and a management procedure (MP) which includes an assessment method and a harvest control rule (HCR) (Kell et al., 2005; Dichmont et al., 2006; Rademeyer et al., 2007). The OM is assumed to represent "true" resource dynamics and is the basis for generating assessment and projection data. Next, the assessment model is fitted to the generated data, and finally, a projection of catch and biomass development is performed. Performance statistics are calculated, allowing for conclusions to be drawn on the usefulness of a given management method in light of the precautionary approach and fishery benefits. Other important processes, such as decision making and fishery adaptation, may also be included in the simulations. In most cases, MSEs have been used to test the

* Tel.: +48 58 73 56 267; fax: +48 58 73 56 110.
E-mail address: horbowy@mir.gdynia.pl

0165-7836/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.fishres.2010.12.015

performances of different management strategies using a given assessment method. Management procedures employing different assessment methods (e.g., age-structured models, production models or difference models) have rarely been contrasted. An example of such an application comes from Punt (1993), who compared the application of a surplus production model and an *ad hoc* tuned VPA (virtual population analysis) for the management of the Cape hake stock off the western coast of South Africa, and Dichmont et al. (2006) who tested three different assessment methods in management procedures for the Australia's Northern prawn fishery.

In the present study, the MSE approach was used to evaluate the performance of management procedures utilising three assessment methods and four HCRs. The assessment methods tested were: the XSA (Shepherd, 1999), the Schaefer production model (Schaefer, 1954), and the difference model (Horbowy, 1992). This report evaluates both the ability of the stock assessment methods to estimate biomass using historical data as well as the performance of these methods in achieving stock management objectives (e.g., high stable catches) when they are combined with harvest control rules. In terms of biological parameters, the stock generated for the OM resembled the eastern Baltic cod stock. Model performances were tested on generated stock that was similar to Baltic cod because of the limited progress in consistent age interpretation for cod, despite the fact that work on ageing consistency has been conducted since the early 1990s (ICES, 2006a). The inconsistency in ageing may affect stock assessment and prediction; thus, it is of value to test whether simpler assessment methods can perform satisfactorily in stock management.

2. Methods

The terms operating model (OM), management procedure (MP), harvest control rule (HCR), performance statistics, process error, and observation error are used as defined by Rademeyer et al. (2007).

2.1. Operating model

A model of a stock with specific dynamics was generated for a period of 20 years (e.g., 2005–2024) using two classical equations of stock dynamics: the exponential decay of cohort numbers and the Baranov catch equation (Appendix A.1). Stochasticity (process error) was introduced into the generated values by adding random lognormal error to the following values: initial stock numbers, recruitment depending on the spawning stock biomass (SSB), and fishing mortality (F). The generated stock resembled eastern Baltic cod stock ($M = 0.2$, maximum weight of approximately 10,000 g, begins maturing at age two, and most fish are mature at age four and older) and included the following characteristics:

- Initial stock numbers, weight-at-age in the catch and stock, and maturity and selectivity-at-age, as estimated in ICES (2006b).
- A hockey-stick sub-model for the expected value of recruitment that is dependent on the SSB (i.e., recruitment increases linearly with biomass to a specific spawning stock biomass level, and the next recruitment is constant).

In some simulations, the Ricker (1954) sub-model for recruitment dependence on biomass was used to test the robustness of the management procedures for different stock-recruitment relationships in the OM.

Fishing mortality (F) from 2005–2024 was assumed to be proportional to fishing effort (f), with catchability that was dependent on age but constant over time. The dynamics of fishing effort in the generated stock were determined according to the four scenarios presented in Fig. 1. These scenarios include f increasing or decreasing by 5% per year (scenarios 1 and 2), f increasing by 7% per year in first 10 years and then fluctuating (scenario 3), and f equivalent to F as estimated for Baltic cod from 1986 to 2005 (ICES, 2006b), scaled to 75% of the average F (scenario 4). These fishing effort options were selected to resemble the typical patterns of f in different fisheries. In options 1–3, the fishing effort value for first year was equivalent to an F of 0.4.

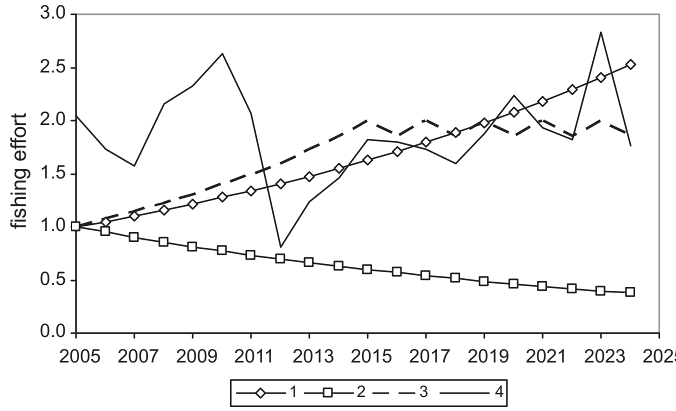

Fig. 1. The four fishing effort options considered in the "true" stock in the OM for 2005–2024.

2.2. Management procedure

The typical ICES procedure for stock assessment and advice on catch quotas was simulated. In this procedure an assessment performed in year y uses the assessment data for years up to $y-1$, so that the stock estimates for the beginning of year y can be obtained. Then, an assumption for the catch in year y is made (e.g., equal to the total allowable catch (TAC) set for that year), and the stock at the beginning of year $y+1$ can be projected. Next, catches for year $y+1$and stock size at the beginning of year $y+2$ are projected using given HCR and other selected options for fishing mortality.

Thus, to simulate ICES procedure, the data for stock assessment (total catch, catch-at-age in numbers, fishing effort, survey indices of stock size-at-age, including recruitment and mean weight in the stock) were drawn from the "true" stock (i.e., from the OM), and the assessment model was fitted. These assessment data were assumed to be distributed log-normally, with averages equal to those from the "true" stock and with a specified sampling variance (observation error). Two options for sampling variance were assumed, representing low and high levels of data collection error (Table 1).

After fitting the stock assessment model to the generated data, projections of stock size and catches were performed using specific harvest control rules, and the projected catches were assumed to be the catch (TAC) for the following year. These catches were then applied to the "true" stock. The assessment data from the projected (realized) catches and the "true" stock sizes were drawn, and new assessments and projections were performed. This sequence of consecutive assessments and projections was repeated for the next seven simulated years, i.e., 2025–2031 (Fig. 2). The time span of simulations was constrained to seven years, as the management plans are usually re-evaluated after a few years. The weight-at-age, maturity, and natural mortality were kept constant in the OM for 2025–2031 and equal to the averages of the values used for 2005–2024. Sampling from generated data, assessing and reassessing the stock, and projecting the catches were usually performed with 200 replications. Simulations were conducted in R using some elements of FLR including the FLXSA package (Kell et al., 2007; http://flr-project.org/).

Table 1
The two options for standard error (s.e.) used to sample assessment data from operating models (the s.e. of effort and mean weight were kept constant).

s.e.	Yield	Catch at age numbers	Survey biomass	Survey at age numbers	Effort	Mean weight
Low	0.10	0.20	0.20	0.30	0.05	0.10
High	0.20	0.40	0.30	0.45	0.05	0.10

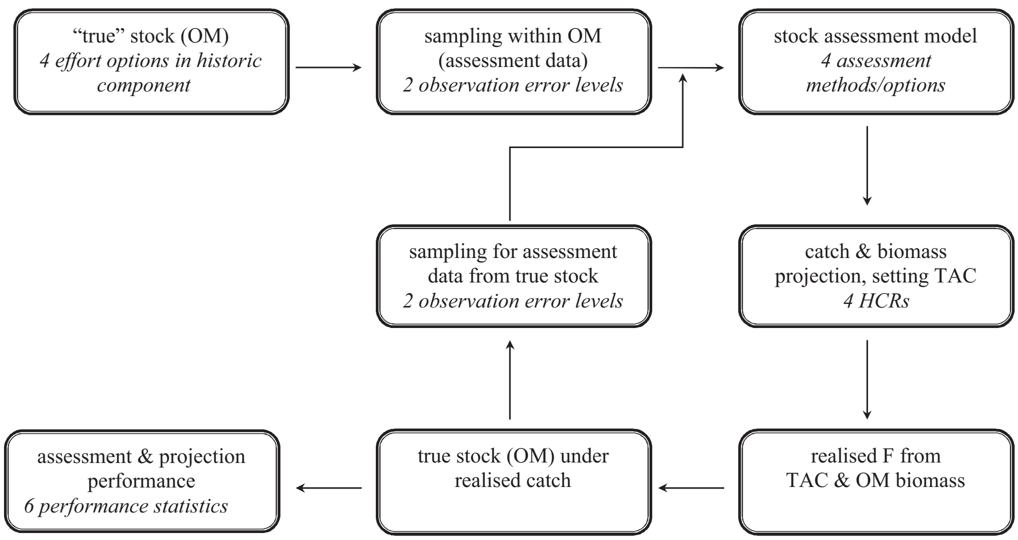

Fig. 2. The scheme of the simulations; the loop is repeated for seven years, 2025–2031 (OM = operating model, TAC = total allowable catch, HCR = harvest control rule).

2.2.1. *Assessment models*

The Schaefer (1954) model used is described by:

$$\frac{dB}{dt} = rB(B_\infty - B) - qf(t)B, \tag{1}$$

while the difference model (Horbowy, 1992) was used in the form

$$B(t+1) = \exp[Hw(t)^{-1/3} - qf(t) - M - k]B(t) + R(t+1), \tag{2}$$

$$R(t+1) = aRindex(t+1), \tag{3}$$

where B = biomass, t = time, q = catchability, B_∞ = carrying capacity, r = intrinsic growth rate, M = natural mortality, H and k = individual growth parameters (anabolism and catabolism coefficients, respectively), w = average weight in the stock, R = recruitment, $Rindex$ = recruitment index from the survey, and a = scaling coefficient. The Schaefer model was parameterized as in Fletcher (1978) so that the maximum sustainable catch (m) was a model parameter. The parameters estimated within the Schaefer model were the initial biomass (B_0), B_∞, q, and m, while within the Horbowy (1992) model, they were B_0, q, and a. For the difference model, the M was assumed to be known and was the same as in the OM. The growth parameters H and k were assumed to be estimated outside the model (e.g., on the basis of one year of data if growth changes do not show a clear trend); other input data were the recruitment index and average weight in the stock. The Schaefer and difference models were fitted to sampled catches by minimizing the sum of the squared residuals between the "observed" and modelled catches in weight (C):

$$SS(param) = \sum_t (\ln C_t^{obs} - \ln C_t)^2, \tag{4}$$

where *param* is a vector of unknown parameters.

The XSA method (Shepherd, 1999; implemented as in Darby and Flatman (1994)) is a way of tuning classical VPA to age-structured survey or commercial catch per unit of effort (CPUE) data. At each age of a cohort's life for which the CPUE index is available, the population in the terminal year may be determined from the ratio of the CPUE index and catchability at a given age multiplied by an exponential term showing loses due to subsequent natural mortality and catches. Such estimates of terminal populations are averaged over cohort ages with weights dependent on CPUE variance and final estimates of terminal numbers are derived. The method has an advantage over *ad hoc* tuning methods because it is not as sensitive to errors in CPUE in the terminal year. The XSA was applied with default settings, with the exception that for the s.e. of the mean to which F estimates shrink (see Appendix A for shrinkage explanation), the following two options were used: 0.5 (default shrinkage) or 2 (low shrinkage). A summary of the data used by the considered assessment methods and estimated quantities are presented in Table 2.

Table 2
The data used to run the considered assessment methods and quantities (variables) estimated by these methods.

Data/variable	Assessment model		
	XSA	Schaefer	Difference
	Assessment data		
Catch volume	Yes	Yes	Yes
Catch at age in numbers	Yes	No	No
Weight at age	Yes	No	No[a]
Fishing effort	No	Yes	Yes
Survey data:			
Stock size at age	Yes	No	No
Recruitment index	Yes	No	Yes
Mean weight in the stock	No	No	Yes
	Estimated variables		
Stock biomass	Yes	Yes	Yes
Fishing mortality (F)	Yes	Yes	Yes
F at age	Yes	No	No
Stock size at age	Yes	No	No

[a] Not needed directly but difference model uses growth parameters.

2.2.2. *Harvest control rules*

Four HCRs were considered. In the first (Rule1), fishing mortality, which was used to set the TAC for years 2025–2031, was related to the current biomass and biological reference points (BRPs). Rule1

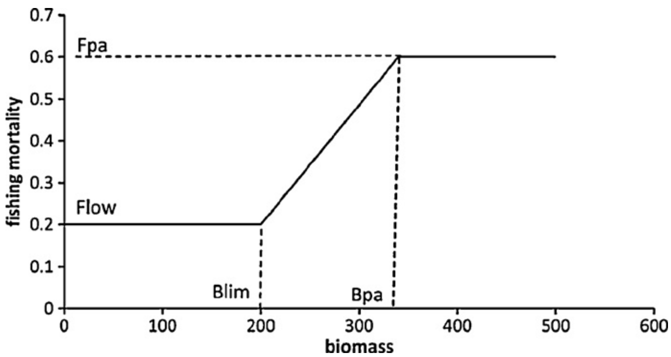

Fig. 3. The dependence of fishing mortality on stock biomass (10^3 tons) in HCR Rule1 for F_{low} = 0.2, F_{pa} = 0.6, B_{lim} = 200,000 t and B_{pa} = 340,000 t.

(Fig. 3) assumed $F = F_{low}$ for SSB < B_{lim}, $F = F_{pa}$ for SSB > B_{pa}, and F linearly decreasing from F_{pa} to F_{low} for B_{lim} < SSB < B_{pa}, where B_{lim}, B_{pa}, F_{lim}, F_{pa} are BRPs (see Appendix A for definitions of BRPs) and F_{low} was selected as a low fishing mortality, allowed when the stock size was below B_{lim}. If the rule assumed no fishing at SSB < B_{lim}, then F_{low} was set to 0. The second HCR (Rule2), assumed fishing mortality was decreasing (or increasing) by p% per year until it reached some required level and p was assumed to be 10%, in accordance with the EU management plan for eastern Baltic cod.

The other two rules (Rule1a and Rule2a) were based on Rules 1 and 2, but included TAC constraints. The maximum change in the TAC was constrained to be 15% of the previous year's TAC.

B_{lim} was estimated using a hockey-stick recruitment model as the SSB producing 50% of maximal recruitment (Myers et al., 1994). F_{lim} was determined to be consistent with B_{lim}, and B_{pa} and F_{pa} were derived from B_{lim} and F_{lim}, assuming an assessment standard error of 0.3 (standard ICES procedure, ICES (2009), see Appendix A). The values of BRPs in the OM were as follows: B_{lim} = 200,000 t; B_{pa} = 328,000 t; F_{lim} = 0.6; F_{pa} = 0.37; and F_{low} = 0.2. For a given assessment procedure and Rule1 or Rule1a, in setting the TAC, the BRPs were rescaled from the corresponding BRPs in the OM following the regression of the assessment biomass and F estimates versus the biomass and F values generated in the OM. The reason for this rescaling was that BRPs are usually estimated based on the F and SSB values obtained in the stock assessment.

2.2.3. Performance statistics

To evaluate how the different models performed in stock assessments and projections, the following performance statistics were considered for each assessment and projection method:

(1) The root mean square relative error (RMSE) of the SSBs assessed in the first 20 years, i.e., 2005–2024.
(2) The root mean square relative error of the estimated terminal SSBs in the next seven prediction years, i.e., 2025–2031.
(3) The probability of the resultant ("true") SSB being above the "true" reference point B_{lim} in each of seven years of the projection.
(4) The variability of the resultant ("true") SSB in the seven prediction years.
(5) The average value of the predicted catch (TAC) in the seven prediction years.
(6) The variability of the predicted catch in the seven prediction years.

Variability was evaluated using the average annual variability (AAV) as defined by Rademeyer et al. (2007). The first two performance statistics relate to assessment performance, while the next four relate directly to MP performance.

An attempt was undertaken to rank each model's performance. The best model for a given f option, performance statistic, and sampling variance was given a mark of 3, and the other models were given marks of 2, 1 or 0, depending on their performance ranks. If two or more models performed similarly (performance statistics values differed no more than 10%), then they were both given the average of the mark values. The marks were averaged across all f options, sampling variance options, performance statistics, and HCRs.

3. Results

3.1. Assessment performance

The results pertain to a hockey-stick stock-recruitment relationship, unless indicated otherwise. Examples of the medians of the retrospective estimates of stock size made with the Schaefer and difference models are presented in Fig. 4a and b, while the retrospective estimates of stock size obtained with the two XSA approaches (default and low shrinkage of fishing mortality to its mean) are shown in Fig. 4c and d. For comparison, the medians of "true" stock size, obtained as a result of realized predicted catches, are also presented. The medians of the estimates were not sensitive to variance in the input data, as data with lower and higher variances produced similar results. The variances of the estimates did not depend markedly on the assumed variance of the input data, as exemplified by the differences in the 10th percentiles of the SSB distributions: those obtained for higher variances of input data were only approximately 5% lower than those produced with lower variances of input data. The root mean square error of the biomass estimates in 2005–2024 for the four fishing effort scenarios is presented in Fig. 5a. In the Schaefer and difference models, the errors were very similar for both options of assessment data variance. For the XSA model assessment, the error was generally markedly lower for input data with a lower variance. In most cases, the XSA model assessment better reproduced the OM stock size than did the Schaefer and difference models. The XSA RMSE was below 0.2, while the error for the Schaefer and difference models was up to 0.6. The difference model usually produced estimates of stock size that were closer to the "true" values than did the Schaefer model; only in the case of a declining f option did the SSB assessment with the Schaefer model have less error than the difference model.

Fig. 5b presents the RMSE for estimates of terminal SSB in consecutive assessments when Rule2 was applied to harvest control. The lowest RMSE was produced by the XSA model (below 0.2), and the model with low shrinkage performed somewhat better than the model with default shrinkage. The difference model generally produced RMSEs in the range of 0.2–0.3, while the errors for the Schaefer model were higher, usually 0.3–0.4. In the case of harvest control with Rule1, the Schaefer and difference models performed similarly (RMSEs of 0.15–0.3), and the XSA model worked well for options with low shrinkage (RMSE of 0.1). However, with a default F shrinkage, the XSA model produced estimates with errors similar to, or even higher than, those of the Schaefer and difference models. For both HCRs with yearly TAC constraints, the RMSEs for estimates of terminal SSB in consecutive assessments were similar to those obtained when an HCR without TAC constraints was applied. The assessment performance of the considered models is summarized in Table 3. On average, the XSA model with low shrinkage produced the lowest assessment error (2.4 points), the XSA model with default shrinkage performed somewhat worse (2.0 points), the difference model performed much worse (1.1 points), and the Schaefer model performed the worst of all of the models (0.5 points). It should be noted that the RMSE of the terminal SSB

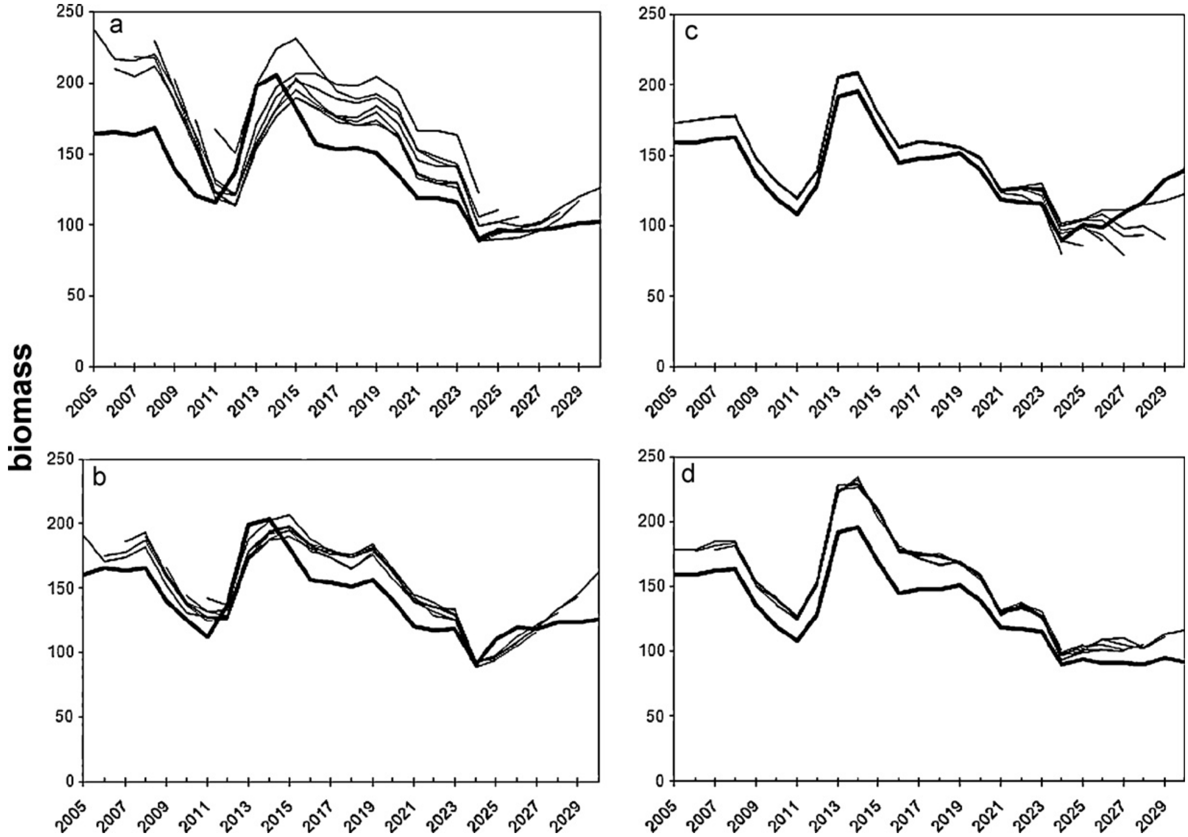

Fig. 4. Examples of the medians of retrospective estimates of stock size (10^3 tons) using the Schaefer model (a), difference model (b), the XSA with default shrinkage (c), and low shrinkage XSA (d) (fishing effort option 4 in 2005–2024, Rule2 used in the prediction, high sampling variance). OM estimates: thick lines; assessment model estimates: thin lines.

also now depends on catches predicted from the HCR, so this comparison is dependent not only on assessment methods, but to some extent, also on the catches predicted from HCRs that differ between the models.

3.2. Management performance

With respect to management performance, for the harvest with Rule1, the Schaefer model performed the worst in terms of the probability of the stock size being above B_{lim} (Fig. 6). The difference model performed much better than the Schaefer model, and the XSA models usually performed the best. In Fig. 6, the probability of the stock size being above B_{lim} when Rule1 is applied with perfect information is shown for comparison. For historical fishing effort option 4, some XSA models produced higher probabilities of the SSB being above B_{lim} than the perfect information scenario, as the assessment procedure largely underestimated stock size, and lower TACs were set, leading to higher stock biomasses. The differences in the models' estimations of the probability of SSB being above B_{lim} were lower with Rule2 than with Rule1. For harvest with Rule2, the Schaefer and XSA models performed similarly, and the difference model performed markedly better. When the TAC constraint was used in the HCR (Rule1a and Rule2a), the difference model performed the best in terms of the probability of SSB exceeding B_{lim}. Both XSA models produced lower probabilities of SSB being above B_{lim} than the difference model, and estimates of that probability for the Schaefer model were between the probabilities for the two XSA approaches. The XSA model with default shrinkage performed better than the low shrinkage approach. As expected, the TAC-constrained Rule1a produced a lower probability of SSB being above B_{lim} than the TAC-unconstrained Rule1 (Fig. 6).

The final performance statistics were the comparisons of the average catches achieved with management by the Schaefer, difference, and XSA models and the variability of catches and SSBs. In the case of harvesting with Rule1, the difference model usually produced slightly higher catches than the Schaefer model (Fig. 7), but catches in the Schaefer model were less variable. The average catches obtained with the XSA model with low shrinkage were lower than the Schaefer and difference model catches but higher than the catches for the XSA model with default shrinkage. For Rule2, the difference model produced the highest catches, and the XSA model with default shrinkage produced the lowest catches, but catches resulting from the Schaefer model and from the XSA model with low shrinkage were similar. The highest variability of catches was from the XSA model with default shrinkage for both Rule1 and Rule2. The XSA model with low shrinkage had a somewhat higher variability of catches than the difference and Schaefer models.

For harvest with Rule1a (Fig. 7), the average catches obtained from the Schaefer and difference model were the highest and similar, while the catches from the XSA model with default shrinkage were the lowest. The differences among catches were smaller than in the case of Rule1 catches, which was an effect of the TAC constraint. For Rule2a the catches obtained from the difference and XSA models with low shrinkage were higher than the catches from the Schaefer model and default shrinkage XSA. The variability of catches for rules with TAC constraints showed similar variability patterns as catches for Rule1 and Rule2, but the differences between AAVs were smaller.

When Rule1 was applied to harvest control, the variability of the biomass in the difference model was somewhat smaller than in the Schaefer model. The variability of SSB in the XSA models was higher or lower than in the difference and Schaefer models

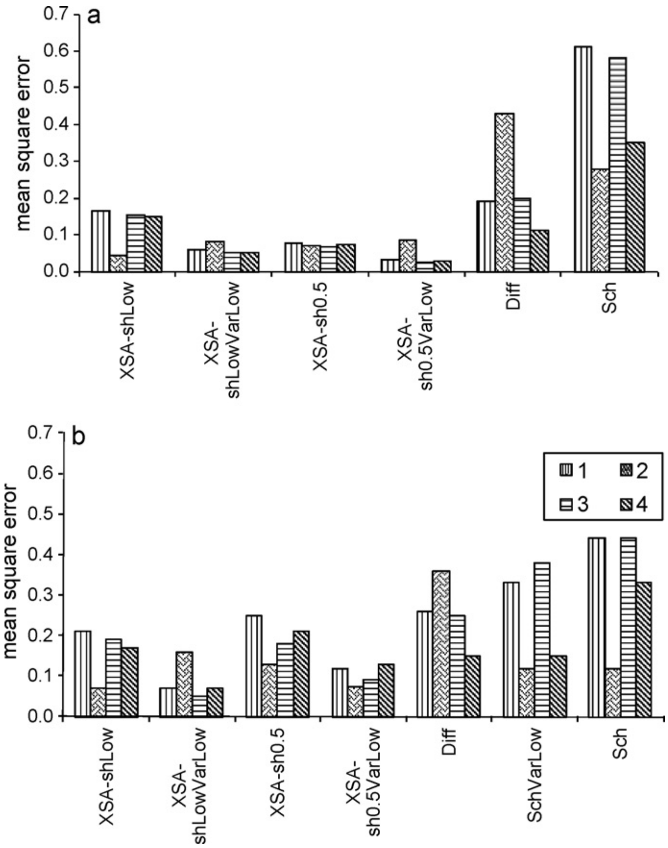

Fig. 5. The root mean square error (RMSE) of the biomass estimates in 2005–2024 (a) and the RMSE of the estimates of the terminal SSB in consecutive assessments, 2025–2031 (b) for the XSA (default and low shrinkage, sh stands for shrinkage), Schaefer (Sch), and difference (Diff) models. Rule2 was used in the predictions. Individual bars within a given model represent four options for the fishing effort. The results for a low variance of input data (VarLow) are shown only if they are markedly different from the results for high variance.

depending on the *f* option and shrinkage used. For Rule2, the lowest variability was obtained with the difference model. In the case of harvesting with Rule1a, the lowest variability was obtained from the low shrinkage XSA model, variability in the Schaefer model was higher, and the variability in the difference model was the highest (Fig. 8). In general, for HCRs with a TAC constraint, the average biomass variability for the Schaefer model was the lowest, while the variability of biomass for both XSA models and the difference model were somewhat higher but similar.

The management performances of the considered models are summarized in Table 3. The difference and Schaefer models performed the best (average of 1.9 and 1.7 points, respectively), the XSA model with low shrinkage performed worse (1.4 points), and the XSA model with standard shrinkage performed the worst (1.0 points). The performance of the XSA model with default shrinkage increased somewhat when changes in F were strongly restricted through the HCR and TAC constraint.

3.3. Effect of a Ricker stock-recruitment relationship on performance of methods

HCR Rule1a was also tested within an operating model with a Ricker stock-recruitment relationship. In this case, the performances of the methods did not differ much in qualitative terms from the performance in the OM with hockey-stick recruitment in relation to the SSB and TAC constraint in the HCR. In the case of the assessment procedure, the XSA model with low shrinkage performed the best (2.3 points), and the XSA model with default shrinkage performed the second-best (1.5 points). However, in the OM with a Ricker stock-recruitment relationship, the Schaefer model performed much better than in the OM with a hockey-stick stock-recruitment relationship (1.4 points) and only slightly differed from the default shrinkage XSA model. The difference model performed the worst (0.9 points). With regard to the management procedure, all models performed similarly to those with the hockey-stick recruitment (1.5, 1.3, 1.3, and 1.6 points for the XSA with low shrinkage, XSA with default shrinkage, difference, and Schaefer models, respectively).

Table 3
Comparison of XSA, difference, and Schaefer model performances by HCR and performance statistics. The averages are calculated for the assessment procedure (AP), management procedure (MP), all performance statistics, and over all HCRs. (best performance: 3 points; poorest performance: 0 points).

Performance statistics	Rule1	Rule2	Rule1a	Rule2a	Mean
XSA-low shrinkage					
RMSE-assessment	2.1	2.1	2.1	2.1	2.1
RMSE-terminal SSB	2.9	2.8	3.0	1.9	2.6
Mean AP	2.5	2.4	2.6	2.0	2.4
Above Blim	1.8	1.1	0.5	1.1	1.1
Variability of SSB	1.8	0.6	2.4	0.6	1.3
Mean catch	1.3	1.5	1.5	2.4	1.7
Variability of catch	1.9	1.0	1.4	1.0	1.3
Mean MP	1.7	1.1	1.4	1.3	1.4
Overall mean	1.9	1.5	1.8	1.5	1.7
XSA-default shrinkage					
RMSE-assessment	2.8	2.8	2.8	2.8	2.8
RMSE-terminal SSB	0.0	1.9	0.6	2.4	1.2
Mean AP	1.4	2.3	1.7	2.6	2.0
Above B_{lim}	2.3	1.6	1.9	1.5	1.8
Variability of SSB	0.0	1.0	1.1	1.6	0.9
Mean catch	0.8	0.4	0.8	0.6	0.6
Variability of catch	0.1	0.5	1.4	1.0	0.8
Mean MP	0.8	0.9	1.3	1.2	1.0
Overall mean	1.0	1.4	1.4	1.6	1.4
Difference model					
RMSE-assessment	0.9	0.9	0.9	0.9	0.9
RMSE-terminal SSB	1.8	0.9	1.6	1.3	1.4
Mean AP	1.3	0.9	1.3	1.1	1.1
Above B_{lim}	1.5	2.1	2.4	2.3	2.1
Variability of SSB	2.5	2.9	0.5	2.4	2.1
Mean catch	2.0	2.6	1.8	2.0	2.1
Variability of catch	1.3	2.0	1.4	1.5	1.5
Mean MP	1.8	2.4	1.5	2.0	1.9
Overall mean	1.6	1.9	1.4	1.7	1.7
Schaefer model					
RMSE-assessment	0.3	0.3	0.3	0.3	0.3
RMSE-terminal SSB	1.4	0.5	0.8	0.5	0.8
Mean AP	0.8	0.4	0.5	0.4	0.5
Above B_{lim}	0.5	1.1	1.1	1.1	1.0
Variability of SSB	1.8	1.5	2.0	1.4	1.7
Mean catch	2.0	1.5	2.0	1.0	1.6
Variability of catch	2.8	2.5	1.9	2.5	2.4
Mean MP	1.8	1.7	1.7	1.5	1.7
Overall mean	1.4	1.2	1.3	1.1	1.3

4. Discussion

4.1. Assessment performance

For the assessment procedure, both XSA models performed much better than the difference and Schaefer models, and the XSA model with low shrinkage performed the best. This result is not surprising, as the XSA model primarily applies stock dynamics equations, which were used for generating stock in the OM (equation of the exponential decay of stock in numbers and Baranov catch equation). The assessment with the difference model was much closer to “true” values than the Schaefer model. The difference model included an explicit term for recruitment, which made it eas-

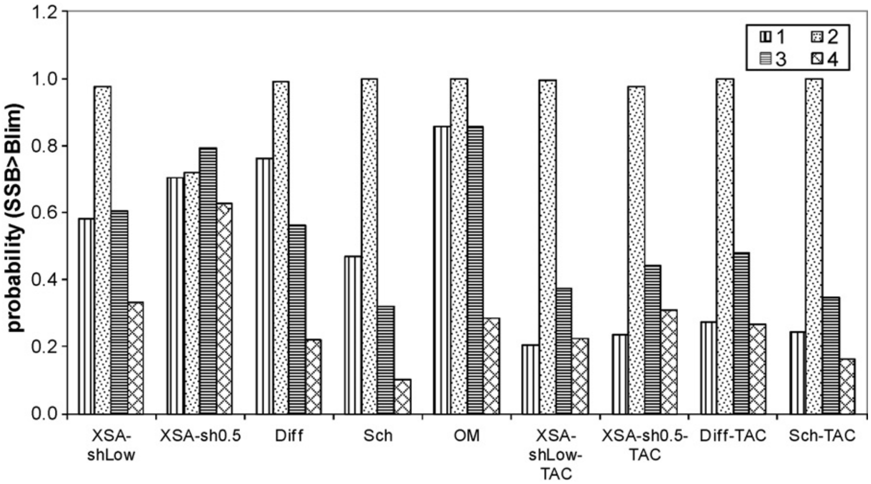

Fig. 6. The probability of the stock size being above B_{lim} for the XSA (default and low shrinkage, sh stands for shrinkage), Schaefer (Sch), and difference (Diff) models. Rule1 and Rule1a (labeled TAC) were used in the predictions, and the results refer to a high variance of input data. Individual bars within a given model represent four options for the fishing effort. The OM bar indicates the probability of the stock size being above B_{lim} when Rule1 is applied with perfect information on stock size.

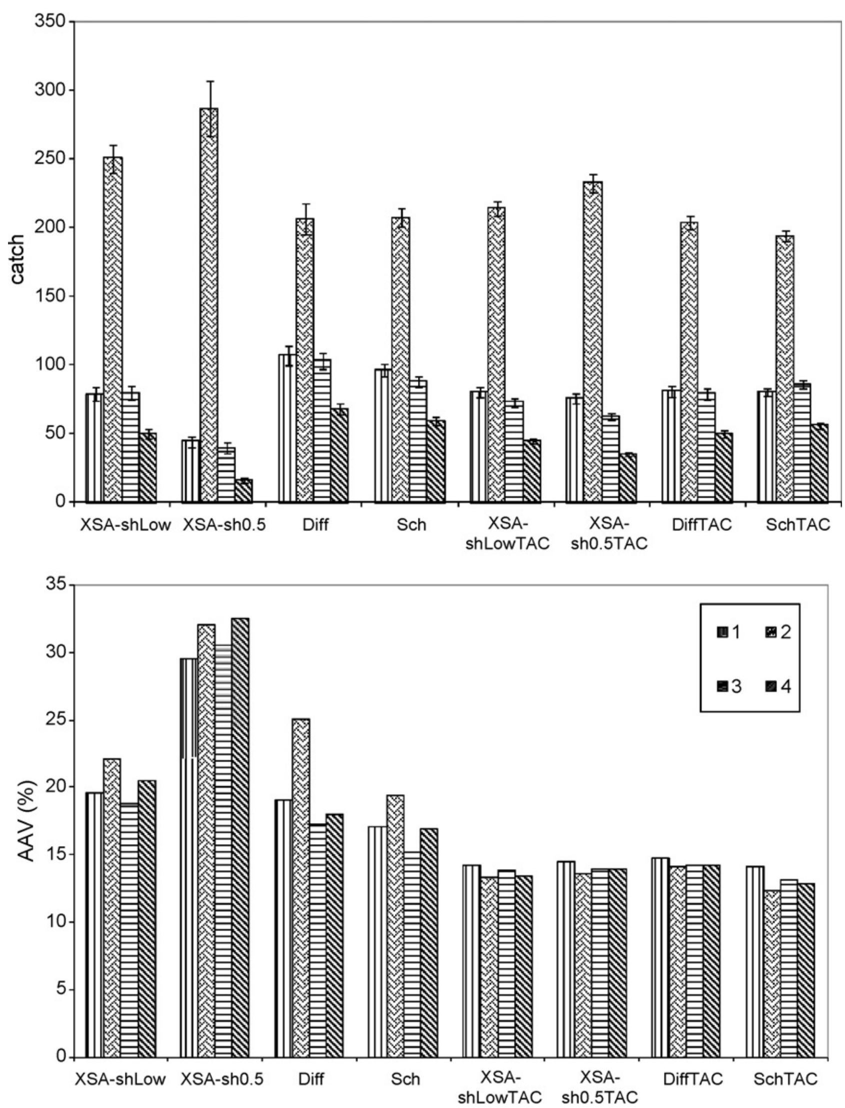

Fig. 7. The mean catch (10^3 tons, whiskers present standard deviation) and its annual average variability (AAV) in 2025–2031 for the XSA (default and low shrinkage, sh stands for shrinkage), Schaefer (Sch), and difference (Diff) models. Rule1 and Rule1a were used in the predictions, and the results refer to a high variance of the input data. Individual bars within a given model represent four options for the fishing effort.

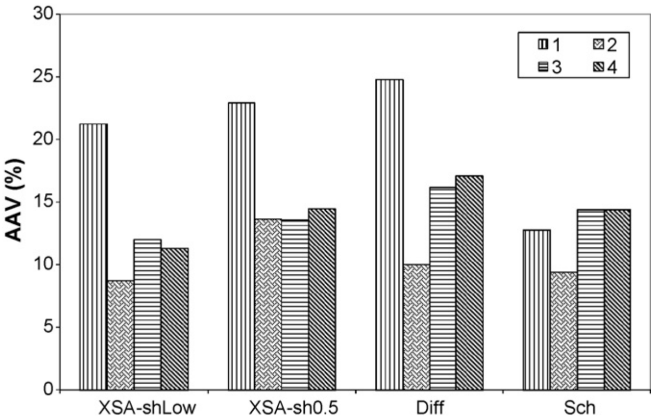

Fig. 8. The annual average variability (AAV) of the SSB in 2025–2031 for the XSA (default and low shrinkage, sh stands for shrinkage), Schaefer (Sch), and difference (Diff) models. Rule1a (TAC constrained) was used in the predictions, and the results refer to a high variance of input data. Individual bars within a given model represent four options for the fishing effort.

ier to reflect stock dynamics compared to the Schaefer model. The assessment errors in the Schaefer model decreased substantially when the Ricker stock-recruitment relationship was implemented in the OM and Rule1a was used in the prediction. Under those conditions, the Schaefer assessment (in terms of RMSE) was only slightly worse than the XSA model assessment with default shrinkage.

4.2. Management performance

The simulations showed that the simpler models might perform well in management procedures, even if they produced high assessment errors that were much larger than those of the XSA models. On average, the difference model performed the best with respect to MP, and the Schaefer model performed slightly worse. Both XSA models performed worse, and the XSA with default shrinkage performed, in all but one case (Rule1a, hockey-stick recruitment), worse than any other model. The XSA model with default shrinkage performed the worst for Rule1 because it fit the stock numbers in terminal years (survivors) as a combination of survivors derived from the survey signal and survivors derived from the average *F* as estimated for the previous five years. The application of Rule1 in the MP resulted in a large change in *F* (from 0.7–1.0 to approximately 0.2 for fishing effort scenarios 1, 3, and 4 in first year of HCR application). In this case, the XSA model with default shrinkage led to a higher bias in *F* estimates for the terminal year, producing an overestimation of terminal fishing mortality and an underestimation of terminal SSB when *F* decreased (the opposite happened when *F* increased). Therefore, the MP using the XSA model with default shrinkage and Rule1 in the prediction led to the highest "true" SSB and was the most precautionary of the considered models. In the case of the other HCRs, the change in *F* was more strongly limited by the HCR itself (Rule2), by the TAC constraint (Rule1a) or by both (Rule2a), so the bias in *F* estimates due to default shrinkage when the trend in *F* was observed was lower. Consequently, the XSA model with default shrinkage performed better for these rules than for Rule1. A similar result was obtained by Bastardie et al. (2010) in their evaluation of a management plan for Baltic cod. They concluded that the plan would be precautionary if the current XSA model settings (especially the default shrinkage option) were kept in subsequent assessments. However, the plan consisted of a gradual decline in fishing mortality until the target *F* value was reached. In this type of situation, the XSA assessment with default shrinkage overestimated terminal *F*, underestimated SSB, and produced a lower TAC than the XSA model with low shrinkage. In light of the simulations described in this study, the use of shrinkage (very popular in ICES assessment working groups) should always be considered carefully. If there are strong changes in recent *F* values, shrinkage may provide greatly biased estimates of terminal fishing mortality and stock size.

The difference model performed the best for harvesting with Rule2 and Rule2a and worse for Rule1a. In most cases, however, it performed better than the Schaefer model. The overall good performance of the Schaefer model in MP was somewhat surprising. However, the Schaefer model was the least precautionary of the models and produced lower probabilities of the "true" biomass being above B_{lim} than most other models, especially when Rule1 was applied. The advantage of the Schaefer model is the direct provision of estimates of *m* (maximum sustainable catch) and *F* at *m*.

The comparison of models performance, summarized in Table 3, assumed equal weighting for all performance statistics, when the averages for assessment procedures or management procedures were calculated. However, applying other weighting of performance statistics could be considered depending on management goals, which could result in the ranking of the methods being different.

The effect of different recruitment functions on model performance was limited. The main effect was a much better fit of the Schaefer model to the assessment data for a Ricker stock-recruitment relationship than in the case of a hockey-stick model. Thus, the performances of the production and difference models were robust for both considered stock-recruitment relationships. The Beverton and Holt recruitment model was not considered because it is similar to the applied hockey-stick relationship.

A somewhat better performance by the difference model compared to the Schaefer model could be expected because the difference model uses some age-structured information (i.e., a recruitment index [or recruitment sub-model], mean weight in a stock, and growth parameters). However, Horbowy (1992) suggested that the model is not very sensitive to assumptions of mean weight if recruitment variability is moderate. In addition, to estimate the mean weight in the stock, the age structure is not required, and growth parameters need to be estimated only once when growth does not undergo large variation.

The applied Rule1 and Rule1a are based on F_{pa} as a target fishing mortality level. ICES is now moving toward an MSY approach following the Johannesburg Declaration of achieving F_{msy} by 2015. The generic HCR using that approach is functionally the same as Rule1, with the only difference being that F_{pa} is replaced by F_{msy} (ICES, 2010). Thus, it is expected that the method performance would not change markedly if the F_{msy} approach was tested instead of using F_{pa} as a target. In addition, as F_{msy} is lower than F_{pa}, reaching the F_{pa} by a given MP means that F_{msy} could also be reached.

4.3. Limitations of the performed analyses

The BRPs applied in the simulations used information on BRPs from the OM. Ideally, the reference points should be estimated based on stock assessment results. However, as B_{lim} is usually defined as the biomass below which recruitment is "impaired" (ICES, 2010), the stock recruitment relationship is required to determine B_{lim}. Of the considered assessment methods, only XSA provides estimates of both recruitment and stock size; the difference model provides recruitment only as a scaled survey index of recruitment, and the Schaefer model does not provide recruitment estimates. Therefore, it appeared to be difficult to apply a consistent means of estimating B_{lim} based on the considered three assessment methods. As a consequence, the simulation results are relevant to situations when all methods provide consistent estimates of BRPs, or the BRPs were set earlier, independent of the considered assessment methods and were used as values scaled to assessment series.

In the simulations with the XSA model, the age distribution was assumed to be unbiased. While there are serious problems with age determination for some stocks, the potential effects of age bias on model performance were not considered in the present analysis. Reeves (2003) showed that for Baltic cod, bias in age determination may affect both the assessment and the prediction of stock size. Thus, the performance of the XSA model with biased ageing would probably be worse than the performance reported in the present study.

4.4. Comparison of obtained results with those reported in literature

Punt (1993) compared the Schaefer (1954) and Fox (1970) production models and an *ad hoc* tuned VPA for the Cape hake stock off the western coast of South Africa. The simulations indicated that in general, the production models performed better than an *ad hoc* tuned VPA. These findings were similar to the results obtained in the present study, where the production model performed similarly or better than the XSA models. Richards and Schnute (1998) compared the performance of a simple model, in which the age structure was represented by mean age only, with complex age-structured models for Pacific Ocean perch. They showed that the biomass estimates by both models were similar and suggested that simpler models may be useful when limited data are available.

Punt (1995) evaluated the performance of management procedures based on the Schaefer production model and found that the model performed satisfactorily when recruitment variability was low and stock productivity was high. Such conditions are met in the stock generated within the OM in the present study.

Some results of Dichmont et al. (2006) showed that in an assessment context, the difference model may perform better than a stock-production model. Furthermore, it was demonstrated that the difference model used to manage the stock did not perform better than simple regression approach, which related log-catch rate to time. These results are qualitatively similar to those obtained in the present study.

Dorner et al. (2009) evaluated management models for Pacific salmon. They used several assessment and prediction models and found that although complex models performed better in some cases, their benefits were small. The models that exhibited smaller assessment or forecasting errors were not necessary better in a management context. Similar results were obtained in the present study: the XSA model produced much lower assessment errors than the difference and Schaefer models, but it did not appear to be better in terms of MP.

Acknowledgements

Part of the present work was conducted with funding support provided by the EU under contract EFIMAS, Project no.: SSP8-CT-2003-502516. The comments and suggestions of three anonymous reviewers are appreciated and helped to improve the final version of the manuscript.

Appendix A.

A.1. Operating model equations

The stock and catch numbers at age were modelled using the exponential decay and Baranov catch equations, respectively:

$$N_{a+1,y+1} = N_{a,y} e^{-(F_{a,y}+M)},$$

$$C_{a,y} = \frac{F_{a,y}}{F_{a,y}+M} N_{a,y}(1-e^{-F_{a,y}-M}),$$

$$F_{a,y} = s_a f_y,$$

where N is stock numbers, C is catch, M is natural mortality, F is fishing mortality, s is selectivity, f is fishing effort, a is age, and y is year. Expected recruitment in the OM is modelled using a hockey-stick sub-model:

$$R = aSSB \quad \text{for SSB} < \text{SSBbreak}$$

and

$$R = R\text{constant} \quad \text{for SSB} \geq \text{SSBbreak},$$

where SSB is spawning stock biomass, SSBbreak is the stock biomass below which recruitment is decreasing, a is the slope of the S-R relationship for SSB < SSBbreak, and Rconstant is the average recruitment for SSB ≥ SSBbreak. The alternative recruitment model tested was obtained from Ricker (1954):

$$R = ar\text{SSB}\exp(-br\text{SSB}), \quad \text{where } ar \text{ and } br \text{ are parameters}$$

Stochasticity (both process error and observation error) was introduced to the MSE through the following formula:

$$Y = \text{Expectation}(Y) \times \exp(norm)$$

where Y was the stock numbers, recruitment, fishing mortality (process error) or assessment data needed to fit the considered assessment models (observation error), *norm* is a normally distributed random variable with a zero mean and a given standard deviation.

A.2. Shrinkage option in the XSA model

The shrinkage to the mean option in the XSA model allows the user to obtain an estimate of the terminal N (survivors) as the weighted mean of two estimators: one resulting from the VPA tuned to the survey data (N_{survey}) and another resulting from the terminal F being the average of *Fs* estimated for a few recent years (N_{Fsh}). In the program (Darby and Flatman, 1994) it is implemented as:

$$\ln N = \frac{\ln N_{\text{survey}}/seN^2 + \ln N_{Fsh}/seFsh^2}{1/seN^2 + 1/seFsh^2}$$

where *seN* and *seFsh* denote the standard error of the N estimate resulting from the survey and standard error of shrinkage mean, respectively. The *seFsh* is defined by the user and its default value is 0.5. The value of *seFsh* should depend on the signal-to-noise ratio in the VPA–survey estimates: the higher the survey noise, the more weight should be given to the average of recent values, making the *seFsh* low.

A.3. Reference points

The reference points used in the study are defined similary to ICES (2010). They are divided into limit reference points (B_{lim} and F_{lim}) and precautionary reference points (B_{pa} and F_{pa}). B_{lim} is the threshold biomass below which recruitment to the stock is expected to be largely reduced (impaired), or stock dynamics are unknown. F_{lim} is the fishing mortality that drives the stock toward B_{lim} if maintained over long period of time. B_{pa} is a precautionary buffer set such that if the estimated biomass is above B_{pa}, then there is high probability that "true" biomass is above B_{lim} (in ICES practical applications, high means approximately 90–95%). F_{pa} is a precautionary buffer set such that if the estimated F is below F_{pa} there is high probability that the "true" F is below F_{lim}. In ICES, some of the B_{pa} and F_{pa} are derived as

$$B_{pa} = B_{lim} \exp(1.645\sigma) \quad \text{and} \quad F_{pa} = F_{lim} \exp(1.645\sigma)$$

where σ is the standard error of the biomass and F estimates and 1.645 is the 95th percentile of the standard normal distribution.

References

Bastardie, F.J., Nielsen, R., Kraus, G., 2010. The eastern Baltic cod fishery: a fleet-based management strategy evaluation framework to assess the cod recovery plan of 2008. ICES J. Mar. Sci. 67, 71–86.

Darby, C.D., Flatman, S., 1994. Virtual Population Analysis: version 3.1 (Windows/DOS) user guide. Ministry of Agriculture, Fisheries and Food. Directorate of Fisheries Research, Lowestoft. Information Technology Series, no. 1.

Deriso, R.B., 1980. Harvesting strategies and parameter estimation for an age-structured model. Can. J. Fish. Aquat. Sci. 37, 268–282.

Dichmont, C.M., Deng, A.R., Punt, A.E., Venables, W., Haddon, M., 2006. Management strategies for short lived species: the case of Australia's Northern Prawn Fishery 2. Choosing appropriate management strategies using input controls. Fish. Res. 82, 221–234.

Dorner, B., Peterman, R.M., Su, Z., 2009. Evaluation of performance of alternative management models of Pacific salmon (*Oncorhynchus* spp.) in the presence of climatic change and outcome uncertainty using Monte Carlo simulations. Can. J. Fish. Aquat. Sci. 66, 2199–2221.

Fletcher, R.I., 1978. On the restructuring of the Pella-Tomlinson system. Fish. Bull. 76, 515–534.

Fox, W.W., 1970. An exponential surplus yield model for optimizing exploited fish populations. Transac. Am. Fish. Soc. 99, 80–88.

Horbowy, J., 1992. The differential alternative to the Deriso difference production model. ICES J. Mar. Sci. 49, 167–174.

ICCAT, 2010. Report for biennial period, 2008–09 PART II (2009) vol. 2, p. 353.

ICES, 2006a. Report of the study group on ageing issues of Baltic Cod. ICES Document CM 2006/BCC: 08.

ICES, 2006b. Report of the Baltic Fisheries Assessment Working Group. ICES Document CM 2006/ACFM: 24.

ICES, 2009. Report of the ICES Advisory Committee. ICES Advice, Books pp. 1–11.

ICES, 2010. Report of the ICES Advisory Committee. ICES Advice, Book 1.

Kell, L.T., O'Brien, C.M., Smith, M.T., Stokes, T.K., Rackham, B.D., 1999. An evaluation of management procedures for implementing a precautionary approach in the ICES context for North Sea plaice (*Pleuronectes platessa* L.). ICES J. Mar. Sci. 56, 834–845.

Kell, L.T., Pastoors, M.A., Scott, R.D., Smith, M.T., van Beek, F.A., O'Brien, C.M., Pilling, G.M., 2005. Evaluation of multiple management objectives for Northeast Atlantic flatfish stocks: sustainability vs. stability of yield. ICES J. Mar. Sci. 62, 1104–1117.

Kell, L.T., Mosqueira, I., Grosjean, P., Fromentin, J.-M., Garcia, D., Hillary, R., Jardim, E., Mardle, S., Pastoors, M.A., Poos, J.J., Scott, F., Scott, R.D., 2007. FLR: an open-source framework for the evaluation and development of management strategies. ICES J. Mar. Sci. 64, 640–646.

Myers, R.A., Rosenberg, A.A., Mace, P.M., Barrowman, N., Restrepo, V.R., 1994. In search of thresholds for recruitment overfishing. ICES J. Mar. Sci. 51, 191–205.

Patterson, K., 1998. Integrated catch-at-age analysis, Version 1.4. Scottish Fisheries Research Report, 38.

Patterson, K., Cook, R., Darby, C., Gavaris, S., Kell, L., Lewy, P., Mesnil, B., 2001. Estimating uncertainty in fish stock assessment and forecasting. Fish Fish. 2, 125–157.

Punt, A.E., 1993. The comparative performance of production-model and *ad hoc* tuned VPA based feedback-control management procedures for stock of Cape hake off the West Coast of South Africa. In Risk evaluation and biological reference points. Can. Spec. Publ. Fisheries Aquat. Sci. 120, 283–299.

Punt, A.E., 1995. The performance of a production-model management procedure. Fish. Res. 21 349–374.

Rademeyer, R.A., Plaganyi, E.E., Butterworth, D.S., 2007. Tips and tricks in designing management procedures. ICES J. Mar. Sci. 64, 618–625.

Reeves, S.A., 2003. A simulation study of age-reading errors for stock assessment and management advice. ICES J. Mar. Sci. 60, 314–328.

Richards, L.J., Schnute, J.T., 1998. Model complexity and catch-age analysis. Can. J. Fish. Aquat. Sci. 55, 949–957.

Ricker, W.E., 1954. Stock and recruitment. J. Fish. Res. Bd. Can. 11, 559–623.

Schaefer, M.B., 1954. Some aspects of the dynamics of populations important to the management of the commercial marine fisheries. Bull. Inter-Am. Trop. Tuna Commun. 1, 25–56.

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