International Council for the Exploration of the Sea

Theme Session on Medium-Term Forecasts in Decision-Making
ICES CM 2000/V:03

Comparison of Uncertainty Estimates in the Short Term Using Real Data

S. Gavaris, K. R. Patterson, C. D. Darby, P. Lewy, B. Mesnil, A. E. Punt, R. M. Cook, L. T. Kell, C. M. O´Brien, V. R. Restrepo, D. W. Skagen, and G. Stefánsson

S. Gavaris: Department of Fisheries & Oceans, 531 Brandy Cove Road, Biological Station, St. Andrews, NB, Canada E5B 2L9 [tel: +1 506 529 5912, fax: +1 506 529 5862 e-mail: GavarisS@mar.dfo-mpo.gc.ca] . K. Patterson: European Commission, Brussels, Belgium. R. Cook: FRS Marine Laboratory, Aberdeen, UK., C. Darby, L. Kell and C. O´Brien: CEFAS Laboratory, Lowestoft, UK., P. Lewy: DIFRES, Charlottenlund, Denmark. B. Mesnil: IFREMER, Nantes, France. A. Punt: CSIRO Marine Research, Hobart, Australia. V. Restrepo: ICCAT, Madrid, Spain. D. Skagen: IMR, Bergen, Norway. G. Stefánsson: MRI & UI, Reykjavik, Iceland.

Abstract

In response to increased interest in the Precautionary Approach, various approaches have been applied to characterize the uncertainty of fisheries assessment projection results. Using three case studies, a comparison of some commonly applied techniques was undertaken to determine if different methods give similar perceptions of uncertainty in the short term, with the same, or very closely similar structural models. The techniques for estimating statistical uncertainty included the delta method, the parametric bootstrap of data, the nonparametric bootstrap of residuals and Bayes. Each method was used to derive cumulative frequency distributions of *SSB* for 1998 and of change in *SSB* for 1998 relative to 1992. These comparisons were contrasted against the sensitivity of uncertainty estimates to fundamental structural assumptions such as separability. Results displayed measurable and often repeatable patterns in differences between methods of estimating uncertainty, suggesting that these differences were peculiar to the methodology and assumptions. The delta method displayed distributions with longer left tails. Results from Bayes and bootstrap percentile methods were similar. Bias adjusted results were more conservative. Often however, differences could be greater when fundamental structural assumptions were altered, indicating that structural relationships must be either clearly established or proper account taken of this model uncertainty.

Introduction

Fisheries management decisions can be conveniently classified into two types. Examples of the first type of decision are characterized by questions like "What is the constant fishing mortality rate corresponding to Maximum Sustainable Yield?" and "What target fishing mortality should be followed in order that the stock should have a less than 5% chance of being under (say) 800,000t in ten years' time?" The former question relates to a steady state situation while the latter is concerned with a transition from a current state to a desired state. An example of the second type of decision is illustrated by the question "What is the catch quota corresponding to (say) 20% exploitation rate?"

The first type of fisheries management decisions are of a strategic nature concerning policies or harvest strategies, in contrast to the second type that are tactical in nature concerning the immediate implementation of regulatory actions. The resulting policies from strategic decisions are often framed in terms of reference points for quantities of interest, e.g. minimum acceptable biomass of (say) 200,000t. Tactical decisions are made in the context of reference points and are therefore dependent on an established harvest strategy. Strategic and tactical decisions are often treated separately because the former require knowledge of production dynamics in order to evaluate alternative options while the latter depend largely on determination of the current state of the resource. Support for estimates of reference points based on modeling production dynamics can be controversial and reference points may be based on practical experience and consensus. This analysis focuses on estimation of uncertainty for making tactical fisheries management decisions and therefore assumes an established harvest strategy with associated reference points.

Until recently, tactical decisions have been based on provision of scientific advice in the form of the "best" point estimate for quantities of interest. For example, the catch quota may have been set at that value corresponding to the point estimate of the projected catch assuming the established fishing mortality reference point. Three factors, the development of statistical methods for estimating stock status, advances in statistical computation techniques permitting more realistic assumptions in complex situations and the emphasis placed on taking uncertainty into account in the, now widely accepted, Precautionary Approach, have stimulated application of risk analyses to fisheries management problems. A diversity of approaches have been used to address the estimation of uncertainty in fisheries (Patterson et al 1999). These approaches involve a broad range of structural and distributional assumptions but also employ different methods for inference. Wade (1999) reviews the strengths and weaknesses of three schools of statistical inference, frequentist, Bayesian and likelihood. The likelihood approach, perhaps presents the most appealing philosophical framework but poses fundamental technical problems and has not been widely applied. This work includes application of frequentist and Bayesian methods but likelihood approaches are not considered.

In this paper, we investigate and compare the perceptions of uncertainty for tactical fisheries management decisions given the same, or very closely similar, structural models. This question is addressed by calculating short-term uncertainty estimates on three real data sets for age structured fishery stock assessments, Eastern Georges Bank haddock, North Sea plaice and Iberian Peninsula sardine.

Methods of Estimating Uncertainty

Both frequentist and Bayesian methods of estimating uncertainty are in common use for making probability statements about interest parameters in fisheries assessment problems.

Probability statements are understood to be based on the confidence (fiducial) distribution of the quantity of interest under repeated sampling for frequentist methods (Efron 1998, Schweder and Hjort 1999) and on the posterior distribution of the quantity of interest for Bayesian methods. Though the interpretation of these probabilities is different, they serve the same purpose and provide the basis of support for decisions under the respective inference paradigms.

Fisheries management interest parameters are often non-linear functions of model parameters and fisheries assessment models are not linear in the model parameters. Estimated confidence distributions or posterior distributions will be displaced for such models. The frequentist notion associated with this characteristic is bias. Adjustment for statistical estimation bias was incorporated for some of the frequentist methods. Bayesian analogues for adjustment of a displacement are not available.

Frequentist

Two generic approaches for obtaining confidence distributions were investigated, delta methods and bootstrap methods. The delta method is a technique for deriving approximate estimates of variance for parameters arising from complex models. These estimates of variance, coupled with some assumption about the sampling distribution of model parameters or of the interest parameter, can be used to construct confidence distributions. The bootstrap is a data based simulation technique that can be used to obtain confidence distributions of interest parameters. This is accomplished by substituting a simple data based estimate for the sampling distribution of a parameter. The parametric bootstrap assumes a parametric form of the distribution but the distribution is characterized by estimates of its defining parameters obtained from the observed data. Non-parametric bootstrap uses the observed data, or residuals about the model fit, directly to define the distribution completely. Results for the bootstrap methods were based on 1,000 replicates.

Delta

The delta method, as used here, involves two steps, initially estimation of statistics for model parameters and secondly translation of uncertainty in the model parameters to the fisheries management interest parameters. Estimation of model parameter covariance was computed in a similar manner, using the common linear approximation (Kennedy and Gentle 1980 p.476), except where the XSA algorithm (Darby and Flatman 1994) was applied. Translating uncertainty of model parameters to risk for fisheries management interest parameters was accomplished either analytically (*an*) or numerically (*num*), as described below. The analytical implementation also made an adjustment for bias.

The Delta method requires further assumptions on which inferences are conditioned, in addition to those made by the assessment model. The analytical Delta makes an assumption about the distribution of the interest parameter, while the numerical Delta makes an assumption about the distribution of model parameters.

The analytical approximation approach was described in Gavaris (1993, 1999). It employs the delta method to estimate the variance of interest parameters from the covariance of the model parameters. An estimate of bias for the model parameters was obtained using Box's (1971) approximation, which requires the assumption that the errors are normally distributed. Bias of interest parameters was derived according to Ratkowsky (1983). Assuming that the interest parameter, η, is distributed according to a Gaussian, the confidence distribution of the bias adjusted interest parameter was approximated as $N(\hat{\eta} - Bias(\hat{\eta}), Var(\hat{\eta}))$. Because the distribution is displaced to adjust for the bias we refer to this as the *an-shift*Delta variant. The increase in variance due to the variance of the bias adjustment was disregarded.

The delta method was implemented numerically using a resampling technique. The covariance of log surviving population numbers was used to draw random samples from a multi-lognormal (Patterson and Melvin 1992). The replicate population numbers were then used to derive replicates of the interest parameter with which a confidence distribution was constructed. A similar resampling technique was used for the XSA implementation, except that the covariances were disregarded, on the assumption that its impact is negligible. The calculation was performed by drawing random samples from independent lognormal distributions defined by the point estimates of the survivors and their standard errors.

Parametric bootstrap

Parametric bootstrap samples were generated by assuming that the indices were distributed according to a lognormal characterized by the estimated mean and variance of the observations (Restrepo et al 1992). The sample replicates were subjected to the entire assessment procedure to obtain replicate estimates of the interest parameter with which a confidence distribution was constructed. Efron (1979) introduced the bootstrap as an automatic way of obtaining better confidence distributions for complex situations. This particular bootstrap technique is referred to as the percentile (*perc*) method.

Nonparametric bootstrap

Ideally, for the nonparametric bootstrap, the observed data would be resampled with replacement to generate sample replicates which could be subjected to the estimation procedure. Smith and Gavaris (1993) employed this approach, but data limitations may complicate routine application. A practical alternative is to resample with replacement from the residuals to the model fit and add these to the predicted values to generate replicate samples (Efron 1993). An example of such an approach in fisheries assessments is provided by Mohn (1993). Because of its reliance on the model fit, this method is referred to as the model conditioned bootstrap, though it is nonparametric because it does not require specification of a parametric distribution for the residuals. When the residuals are not assumed to be homogeneous, the weighted residuals are scaled to the appropriate variance for the respective data before being added to the predicted values. This is straightforward for the indices which are assumed lognormal but presents some complications for the catch at age data when a multinomial is assumed (Annex 1).

As with the parametric bootstrap, the percentile method is simply based on the confidence distribution constructed from the replicates of the interest parameter that are obtained from subjecting the model conditioned sample replicates to the estimation procedure. Efron (1982) introduced an improvement, the bias corrected (*bc*) percentile method, that adjusts for differences between the median of the bootstrap percentile density function and the estimate obtained with the original data sample. Application of the model conditioned bias corrected bootstrap method in fisheries assessment is described in Gavaris and Van Eeckhaute (1998).

Bayes

The posterior distribution of interest parameters were estimated using the Sampling-Importance-Resampling (SIR) algorithm described by Rubin (1987) or Markov Chain Monte Carlo (MCMC) simulation and graphical models described by Gilks et al. (1996). Sampling Importance Resampling uses an importance function of model parameters to obtain importance ratios that can be used as weights in resampling. In Markov Chain Monte Carlo, samples are drawn from required distributions, constructed using Markov chains for a long time, and averaged to approximate expectations. Application in fisheries assessment of the SIR algorithm is described in McAllister et al (1994) and of the MCMC algorithm in Patterson (1999). Because SIR and MCMC algorithms are simply alternative numerical methods for the same purpose and should give similar results when implemented appropriately, we do not distinguish between them. Results for the Bayesian methods were based on either 1,000 draws from the posterior distribution (Bayes1) or 100 draws (Bayes2).

Fisheries Problem

A typical tactical fisheries management question might be "What is the probability that the resulting projected spawning stock biomass will be lower than the established reference for alternative catch quotas?" (Figure 1). In mathematical terms, we wish to characterize $\Pr\{SSB_{proj} \le SSB_{ref} \mid quota\}$. Reference points may be externally prescribed absolutely, e.g. 200,000t, or they may be prescribed by a functional rule and require estimation, e.g. the biomass corresponding to Maximum Sustainable Yield or the estimated biomass in some earlier year. When the reference point is also estimated, the uncertainty in that estimate of the reference point is conveniently incorporated by considering the quantity of interest to be a function of the projected value and the reference value. For example, if we consider the difference between the projected value and the reference point, the mathematical form can be rearranged as $\Pr\{SSB_{proj} - SSB_{ref} \le 0 \mid quota\}$. Now the interest parameter to be estimated becomes $SSB_{proj} - SSB_{ref}$ instead of simply SSB_{proj}.

As indicated above, risks identified with tactical decisions are largely dependent on the uncertainty associated with the estimate of the current stock status. Accordingly, for the purpose of this study, it was sufficient to compare probability statements for the quantity of interest, e.g. *SSB*, in the terminal year of the assessment (Figure 2), thereby avoiding the need to conduct projections and to explicitly consider the objectives and the harvest strategy. It is recognized that the overall risks could be refined by incorporating uncertainty associated with forecast weight at age and forecast exploitation pattern by age. Forecast recruitment is generally not a major concern with short term projections.

As noted, uncertainty estimation is conditioned on structural and error distribution assumptions. The three studies shared some common fundamentals. All three assessments were based on age structured analyses where mortality processes were partitioned into two types, fishing mortality associated with the harvest and natural mortality associated with all other sources of depletion. Mortality dynamics were governed by the relationships

$$N_{a+1,y+1} = N_{a,y} e^{-\left(F_{a,y} + M_{a,y}\right)}$$

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

where *N* is population abundance in numbers, *F* and *M* are instantaneous fishing and natural mortality rates respectively, *C* is catch numbers harvested and *a* and *y* index age and year respectively. In all three studies the natural mortality rate, *M*, was assumed constant over ages and time and was assumed known. Some of the software implementations employed the cohort approximation (Pope 1972) to the catch equation, but this is inconsequential here.

The population analyses were calibrated with indices of abundance. The indices of abundance, which were age specific numbers or age aggregated biomass, were assumed to be linked to the respective population quantity by a constant proportional relationship, referred to as catchability, *q*. For estimation methods requiring specification of a parametric distribution, the residuals of the indices of abundance about the model fit were assumed to be independent and lognormally distributed while non-parametric methods made the assumption that the residuals, on the logarithmic scale, were independent and identically distributed. For the haddock and sardine study, homogeneity was assumed while for the plaice study, homogeneity was achieved by weighting indices according to the relative magnitudes of the mean squared residuals for each index source (fleet).

Though this study was mainly focused on comparison of several methods for estimating uncertainty while restricting the underlying structural models to be the same, the implications of altering some key structural features of the dynamics were considered to a limited extent. It is well recognized that the equations governing mortality dynamics, given above, involve more parameters than can be estimated from typical fishery observations. Consequently, several approaches have been developed to reduce the dimensionality of the parameter space.

Fishing mortality dynamics models can be categorized into those that consider error in the catch at age to be negligible relative to other observation error and those that admit error in the catch at age. Two variants of the former class are prevalent, a Virtual Population Analysis (VPA) model with constraints on the oldest age fishing mortality and a VPA with constraints on the oldest age index catchability. The F-constrained VPA model, designated VPA/F, assumes that the fishing mortality rate for the oldest age is equal to the average fishing mortality rate over specified younger ages in the same year, eliminating the need to estimate abundance for those year-classes. The q-constrained VPA model was not investigated in this study. The separable model, used when admitting error in the catch at age, assumes that fishing mortality can be decomposed into independent year effects and age effects. Two parametric specifications are in common use for the admitted error distribution of the catch at age, lognormal (Deriso et al 1985), and multinomial (Fournier and Archibald 1982)and are designated SEP/L and SEP/M respectively.

The VPA/F assessment models were carried out using various software implementations of the ADAPT adaptive framework (Gavaris 1988). Also, though typically used to implement a q-constrained VPA, the XSA extended survivors algorithm (Shepherd 1999) as implemented in the Lowestoft assessment suite (Darby and Flatman 1994), was modified to mimic the VPA/F model and was applied to the North Sea plaice. Some of the SEP/L assessment models were carried out with the ICA integrated catch analysis software (Patterson and Melvin 1992). Other assessment models were custom designed.

Results

For the three studies, Eastern Georges Bank haddock, North Sea plaice and Iberian Peninsula sardine, we compared probability statements for the spawning stock biomass in 1998, SSB_{1998}, and for the change of spawning stock biomass in 1998 relative to 1992, $(SSB_{1998} - SSB_{1992})/SSB_{1992}$. The three case studies were not subjected to all combinations of the three structural model variants and the six methods of estimating uncertainty. Thus we do not have a complete experimental design. Rather, emphasis was placed on comparing across estimation methods for particular structural models of each case. Table 1 defines the acronyms used for the methods and summarizes the combinations that were analyzed.

We have not faithfully reproduced the assessments for these stocks. For some we have used abbreviated data sets. However, it was considered that these altered cases retained the essential elements of realistic assessment problems for the purpose of comparing methods for estimating uncertainty while simplifying the problems sufficiently to expedite computations.

Models were defined to correspond closely to the assessment, within the scope of each structural model class.

Eastern Georges Bank Haddock

The eastern Georges Bank haddock case was based on the assessment by Gavaris and Van Eeckhaute (1998), however, only the DFO spring survey was used with annual catch at age data for 1986 to 1997.

The VPA/F structural models assumed that the fishing mortality for age 8 was equal to the average fishing mortality, weighted by population number, on ages 4 to 7 in the same year. The SEP/L structural models assumed a common selectivity-at-age pattern through the entire time period, 1986 to 1997, and the selectivity for age 8 was set equal to one. One of the SEP/L analyses, labeled Bayes2, constrained the fishing mortality on the last two ages, 7 and 8, to be equal and did not include age 1. The two Bayes variants employed different prior assumptions for model parameters (Table 2). The *perc*PB calculations used standard errors derived from the sampling variances for each age in each year of the survey.

SSB in 1998

Results are summarized in Table 3 and Figs. 3 - 4. Within the VPA/F structural models, the standard deviation of the distributions were similar for all estimation methods except *perc*PB. Though this summary statistic of dispersion suggests a common perception of spread, examination of the median scaled percentiles revealed some finer differences. While median scaled 25th and 75th percentiles were quite similar, the *an-shift*Delta results had a longer lower tail resulting in a smaller 5th percentile and the *bc*NPB results were somewhat tighter and more asymmetric. The scaled percentiles were very tight and almost symmetric for the *perc*PB method. The mean and median were lower for the *shift*Delta and *bc*NPB, suggesting that the location estimate is affected by non-linearity induced estimation bias. Consequently, the percentiles and the distributions for *shift*Delta and *bc*NPB were both centered to the left of the others. The percentiles and the distributions for *perc*NPB and Bayes were virtually identical. The percentiles for *perc*PB were very tight and the distribution was the least slanted. Confidence statements based on the *perc*PB method would be markedly different. Excluding the *perc*PB results, confidence statements for outer probability levels, i.e. 5th and 95th percentiles, were substantially different with some critical values deviating by almost 10,000t while confidence statements for central probability levels, i.e. 25th and 75th percentiles, were more similar but some critical values still deviated by as much as about 7,000t.

Within the SEP/L structural model, the standard deviation of distributions showed greater differences, though results for *num*Delta and Bayes1 were almost identical. As with the VPA/F results, the *num*Delta median scaled percentiles were smaller reflecting the longer lower tail. The scaled percentiles for *perc*NPB, *bc*NPB and Bayes1 were fairly similar with the exception of the very large value for the 95th percentile for *perc*NPB. The scaled percentiles for Bayes2 were substantially tighter. The percentiles and distribution for the *bc*NPB were shifted to the left of the results for others, again suggesting that estimation bias had some influence. The percentiles and distribution from Bayes2 indicated greater precision than any of the other methods. While critical values for confidence statements at the 25th or 75th probability levels only differed by about 5,000t, those at the 5th and 95th probability levels differed by as much as 37,000t. The differences between the two Bayes variants suggest an important influence of choice of priors on results and/or that the additional constraint and the exclusion of age 1 for the Bayes2 formulation were consequential.

Results from the VPA/F structural models were generally centered about lower *SSB*, displayed smaller standard deviation and had tighter scaled percentiles than those for the SEP/L structural models. The distributions for SEP/L models showed longer upper tails. Confidence statements were somewhat more divergent between structural models than within.

Change in 1998 SSB relative to 1992 SSB

Results are summarized in Table 3 and Figs. 5-6. Within the VPA/F structural models, the patterns in dispersion, location and distributions of relative change in *SSB* mirrored those for *SSB* in 1998 very closely. Notably, even the relative dispersion, scaled to the median, was not too dissimilar.

In contrast, within the SEP/L structural models, the patterns in dispersion, location and distributions of relative change in *SSB* were substantially different than those for *SSB* in 1998. The standard deviation, scaled percentiles, percentiles and distributions for *perc*NPB, *bc*NPB and Bayes1 were virtually identical. The similarity between *perc*NPB and *bc*NPB results suggests that statistical estimation bias was not significant here. The standard deviations and scaled percentiles for *num*Delta and Bayes2 results suggested substantially less precision in the estimates. However the distribution for *num*Delta was centered about a lower value and that for Bayes2 was centered about a higher value than distributions for *perc*NPB, *bc*NPB and Bayes1. Confidence statements derived from *perc*NPB, *bc*NPB and Bayes1 would be very similar while those from *num*Delta and Bayes2 would differ markedly from the others and between themselves. The distributions for the two Bayes variants diverge as probability increases.

SEP/L structural models resulted in lower estimated relative change for *SSB* than the VPA/F structural models. SEP/L models in combination with *perc*NPB, *bc*NPB and Bayes1 estimation methods indicated greater precision for estimated change in *SSB* than the results from VPA/F models while SEP/L models in combination with *num*Delta and Bayes2 indicated the contrary.

North Sea Plaice

The North Sea plaice case is based on the assessment done by the ICES working group (ICES 1998) and uses catch data for ages 0 - 13+ over the years 1988 - 1997, two stock size indices at age from trawl surveys and two stock size indices at age from fishery catch and effort data. Catch at age data is available for years prior to 1988 for stock reconstruction but as there were no stock size indices for these years, they were excluded from analyses done here.

The VPA/F structural models assumed that fishing mortality for age 12 and the 13+ age group were equal to the arithmetic average for ages 10 and 11 in the same year. The SEP/L structural models assumed a common selectivity-at-age pattern through the entire time period, 1988 to 1997, and assumed that selectivity on age 12 and the 13+ age group were equal to the average for ages 9 to 11. As with haddock, the two Bayes variants employed different prior assumptions for model parameters (Table 2). Index observations were weighted according to the inverse of the mean squared residuals for each of the four index sources (fleets) to account for potential heterogeneity. Unlike the textbook approach for bootstrap that is not model conditioned, where standard errors are calculated from the observed data, estimates of standard error from the model fit were used to generate random deviates that were added to the observations to obtain replicates for the *perc*PB estimation method. Estimated model conditioned variance is generally greater than sampling variance, so this implementation is a hybrid.

SSB in 1998

Results are summarized in Table 4 and Figs. 7-8. Within the VPA/F structural models, the standard deviation of the distributions were very similar for all estimation methods. The *perc*PB results were not markedly different here from the other methods, where the variance used to generate replicates was obtained from the model fit. As with haddock, while median scaled 25th and 75th percentiles were quite similar, the *an-shift*Delta results had the smallest 5th percentile and the *bc*NPB results were somewhat tighter and more asymmetric. The mean and median were lower for the *shift*Delta and *bc*NPB, suggesting that the location estimate is affected by non-linearity induced estimation bias. Consequently, the percentiles and the distributions for *shift*Delta and *bc*NPB were both centered to the left of the others. The percentiles and the distributions for *num*Delta, *perc*PB, *perc*NPB and Bayes were virtually alike. Confidence statements for outer probability levels, i.e. 5th and 95th percentiles, were somewhat different with some critical values deviating by about 23,000t while confidence statements for central probability levels, i.e. 25th and 75th percentiles, were more similar but some critical values still deviated by about 16,000t.

Within the SEP/L structural model, the standard deviation of distributions showed greater differences, with the *perc*NPB and *bc*NPB results displaying a substantially larger magnitude. For this case, the *num*Delta median scaled 5th percentiles were not smaller. There was fair deviation among the scaled percentiles for all estimation methods. Once again however, the percentiles and distribution from Bayes2 indicated greater precision than any of the other methods. The largest deviation was between the two Bayes variants, suggesting a dominant influence of choice of priors on results.

Results from the both the VPA/F and SEP/L structural models were generally centered about similar *SSB*. Confidence statements were very divergent within SEP/L structural models, precluding meaningful comparisons between the two structural model classes.

Change in 1998 SSB relative to 1992 SSB

Results are summarized in Table 4 and Figs. 9-10. Within the VPA/F structural model, the patterns of differences between distribution characteristics from the various estimation methods were almost identical to those observed for *SSB*. Within the SEP/L structural model, the *perc*NPB and the *bc*NPB did not display a larger standard deviation, as was observed for *SSB*. The *num*Delta was more characteristic, displaying smaller a smaller median scaled 5th percentile and a distribution with a longer lower tail. The *perc*NPB and the *bc*NPB were in much closer agreement, suggesting that the estimation bias for relative change in *SSB* was not very important here. The Bayes1 distribution was very similar to the *perc*NPB and the *bc*NPB distributions while that for the *num*Delta was in closer agreement with the Bayes 2 results.

The distributions for SEP/L structural models were centered about lower relative change in *SSB* than those of the VPA/F structural models, although there was as much difference within the SEP/L results as between the two model classes. The standard deviation and median scaled percentiles were fairly similar across all methods from both structural models, with the exception perhaps of the Bayes2 results which indicated greater precision.

Iberian Peninsula Sardine

The Iberian Peninsula sardine case is based on the assessment made by ICES (1999) and uses catch for age 0 - 6+ over the years 1977 - 1997, two acoustic surveys with abundance at age for ages 1 - 6+, an egg index of spawning biomass and two CPUE indices of spawning biomass (ICES 1999). The catchability for the egg survey was assumed to be one, i.e. considered to be an absolute index rather than a relative index, in contrast to all other cases where the catchabilities were estimated.

The SEP/L and SEP/M structural models both assumed two selectivity-at-age patterns, one for the period 1986 to 1989 and another for the period 1990 to 1997. Further, the selectivity for the 6+ age group was set to equal one. The VPA/F structural models assumed that the fishing mortality on the 6+ age group was equal to the fishing mortality on age 5.

SSB in 1998

Results are summarized in Table 5 and Figs. 11-13. Within the VPA/F structural model, the standard deviation of the distributions were similar for all estimation methods applied although somewhat smaller for the *bc*NPB method and higher for the Bayes method. Thus these methods provided a fairly comparable perception of dispersion. The scaled percentiles were similar at the 25th and 75th percentiles though the *bc*NPB was tighter and the Bayes was most assymetric. There was more divergence at the 5th and 95th scaled percentiles with the *an-shift*Delta having a smaller 5th percentile and the Bayes results having a larger 95th percentile. The mean and median were lower for the *shift*Delta and *bc*NPB, suggesting that the location estimate is affected by non-linearity induced estimation bias. While the percentiles and distributions for the *an-shift*Delta and *bc*NPB were both centered to the left of the others, the *bc*NPB percentiles were tighter while the *shift*Delta displayed a long left tail. The percentiles and the distributions for *perc*NPB and Bayes were virtually identical with a somewhat longer upper tail for the Bayes results. Confidence statements for outer probability levels, i.e. 5th and 95th percentiles, were substantially different with critical values being separated by over 100,000t while confidence statements for central probability levels, i.e. 25th and 75th percentiles, were more similar but still separated by as much as about 70,000t.

Within the SEP/L structural models, the standard deviation of distributions were similar with *bc*NPB being lowest and *num*Delta being highest. The scaled percentiles were also very similar, though the *num*Delta and Bayes results had smaller 5th percentiles and larger 95th percentiles. The mean and median for the *bc*NPB were lower than the other methods, indicating an effect from bias correction. The percentiles and distributions were not too dis-similar though the results for Bayes were centered about a substantially higher *SSB*. Results for *num*Delta displayed a longer lower tail and those for *bc*NPB were centered about a lower *SSB*. As with the VPA/F structural model results, critical levels of confidence statements for outer probability levels differed substantially, by about 85,000t, and differences at central probability levels were still notable, but only by about 60,000t.

Within the SEP/M structural models, the characteristics of the distributions, standard deviation, scaled percentiles, means, medians and percentiles were very similar. Confidence statements for this class of model would be similar and very tight from the three estimation methods applied, *perc*NPB, *bc*NPB and Bayes. Estimation bias did not appear to be significant in this case. The Bayes results display a peculiar step behaviour at the upper tail.

Results from the VPA/F structural models were generally centered about higher *SSB*, displayed larger standard deviation and had wider scaled percentiles than those for the SEP/L and SEP/M structural models. The perception of lower dispersion by the SEP/L and SEP/M models is notable considering that they additionally admit uncertainty in the catch at age data. Except in a few instances, the medians from the VPA/F, and the separable models lie outside or very near to critical points for each others' 90% probability interval, indicating that uncertainty due to choice of structural model is larger than statistical uncertainty conditioned on a particular structural model.

Change in 1998 SSB relative to 1992 SSB

Results are summarized in Table 5 and Fig. 14-16. For all three classes of structural model, VPA/F, SEP/L and SEP/M, the patterns in dispersion, location, percentiles and shape of distributions for relative change in *SSB* mirrored those for *SSB* in 1998 very closely.

Results for the VPA/F structural models were generally centered about highest relative change in *SSB*, while results for SEP/L were lowest and those for SEP/M were intermediate. The VPA/F structural models also displayed larger standard deviation and had wider scaled percentiles than those for the SEP/L and SEP/M structural models, but the differences were not as marked as for *SSB*. In contrast to the pattern for *SSB*, for relative change in *SSB* the medians from the VPA/F, and the separable models generally lie within each others' 90% probability interval. Indeed, the distributions for Bayes-SEP/M is very similar to that for *bc*NPB-VPA/F at lower probability levels, diverging somewhat at higher probability levels.

Discussion

Delta, bootstrap and Bayesian methods for making probabilistic inferences about fisheries management interest parameters are in common use. Almost invariably, the choice of estimation method is not discussed. Quite often, the estimation method is selected on the basis of ease with which it can handle particular structural conditioning choices. However, this is not a particularly compelling rational as techniques have been developed for all these estimation methods to handle most structural conditioning situations. It is pertinent therefore to ask if prevalent variants of these estimation methods result in similar inferences when the same structural models are used.

The results demonstrate that there can be differences in the characteristics of confidence and posterior distributions obtained with the different estimation methods, in both location (central tendency) and dispersion (spread), even when the same structural model is used. The magnitude of the differences can be substantial. For example, the difference in the medians of the distributions ranged as high as 20% of the value of the smallest median in most structural models. Differences were often even greater at the tails of the distributions than at central probability levels.

Some regular patterns could be detected in the differences, although these patterns were not faithful for all cases. The patterns appeared to be more consistent and predictable for the VPA/F structural model. Bias adjusted distributions, i.e. *an-shift*Delta and *bc*NPB, were displaced towards lower *SSB* and lower relative change in *SSB* relative to other methods. This suggests a measurable and systematic effect of estimation bias, though less for rellative change in *SSB*. Both Delta variants tended to have longer lower tails in the distribution, though this effect was not as great for the numerical method. This is an indication that the distributions of the interest parameters are not well approximated by a symmetric Gaussian and that even a lognormal distribution for population survivor abundance may not capture the appropriate degree of asymmetry. Sinclair and Gavaris (1996) also found that results from analytical and numerical delta methods, where the same structural model was used, were largely comparable but there were notable differences for one interest parameter at lower probabilities. In several instances, and most notably with the VPA/F structural models, the Bayes distributions corresponded fairly closely with the *perc*NPB results. One can conclude that the particular priors used in these calculations were not very informative and did not influence results. This is not a general result however, as evidenced by the divergence in distributions when different priors were used with the SEP/L structural models. The divergence between the parametric bootstrap and other methods for haddock needs further investigation. The similarity in results for plaice also merit examination because the variances used to generate the deviates might have been expected to result in greater spread for the parametric bootstrap. The direction of the difference between posterior distributions for *SSB* using Bayes1 and Bayes2 are in agreement with previous investigations on the impact of priors (Walters and Ludwig 1984).

We intentionally investigated an absolute interest parameter, *SSB*, and a relative interest parameter, relative change in *SSB*, to determine if there was a difference in behaviour. It might be presumed that relative quantities can be determined with greater precision. While this may occur, it is not a general phenomenon. Our results indicate that the dispersion for the distributions of relative change in *SSB* was similar to those for *SSB*. Further, for the structural models investigated, the patterns of differences between estimation methods that were observed for *SSB* were closely mimicked by those for relative change in *SSB*. From these results we may infer that the differences between the estimation methods are likely to be manifest in most fisheries management interest parameters. A notable characteristic however, was the diminished differences in the distributions across structural models for relative change in *SSB* compared to absolute magnitude of *SSB*, suggesting that relative measures may be more robust to model choice.

Although not a focus of this study, it is noteworthy that there were great differences in the location (central tendency) of distributions across structural models, often larger than some of the differences between estimation methods within a structural model. Higher or lower medians were not associated with any particular structural model. For example, distributions of *SSB* were centered about higher values with the SEP/L model for haddock, but for sardine, the distributions were centered about higher values with the VPA/F structural model. The variation in dispersion (spread) among distributions seemed more similar between estimation methods within structural models than between structural models. More importantly however, any particular structural model was not associated with lesser or greater precision. For example, standard deviation of the distributions or median scaled inter-percentile ranges of *SSB* were tighter with the SEP/L structural model for sardine, but for haddock, they were tighter with the VPA/F structural model. The very close agreement between distributions for different estimation methods with the SEP/M structural model is a peculiar result. The comparisons are not sufficient to draw any conclusions but this phenomenon is worthy of further study.

Even within a model class and particular estimation method, subtle alterations of structure and assumptions can result in substantial impact on the distribution characteristics of fisheries management interest parameters. This is illustrated with North Sea plaice using the SEP/L structural model and the Bayes method of estimating uncertainty. There were marked differences between distributions for five alternative analyses (Fig. 17). The Bayes1 and Bayes2 analyses were described before. The other analyses are based on Bayes2 but Bayes2 & M includes estimation of M, the power q analysis uses a power relationship for catchability rather than a proportional relationship and the RWF analysis incorporates a random walk for fishing mortality. The random walk model is similar to the separable model but permits stochastic variation from this fixed effects pattern for fishing mortality (Ianelli and Fournier 1998).

It is important to recognize that admitting additional error in the data does not correspond to lower precision for the fisheries management interest parameters. There may be a predisposition to assume that separable models, that admit error in the catch at age, will result in greater uncertainty for interest parameters, but this is not the case. The degree of uncertainty in the interest parameter is more closely associated with the fit of the specific data to a particular structural model. The greater dispersion in the results for haddock with the SEP/L structural model compared to the VPA/F model was probably due to the marked shift in the exploitation pattern at age which occurred in the early 1990s, resulting in a poor fit to a model that specified a common age effect over all years.

It is clear that the choice of structural model has profound effects on inferences about fisheries management interest parameters. Careful consideration should be given to all available diagnostics for determining the most suitable model, consistent with observed data. Techniques, particularly within the Bayesian paradigm, are available to admit more than one structural model (Patterson 1999). These approaches may offer advantages in cases where the data do not strongly favour any particular model. Although greater attention has been given lately to model averaging with frequentist methods (Buckland et al 1997), development of established techniques that can be applied to fishery stock assessment models requires further work. When the data are not informative with respect to model selection, choice of a single model or relative preference among competing models may be based on subjective judgement or expert opinion. Inferences are conditioned on these choices and it should be made clear where subjectivity has an influence versus where observed data are dominant. Proper interpretation of probabilistic inferential statements requires extra care when model indeterminacy is involved.

Though the choice of structural model can have profound impact on the estimation of uncertainty, the regularity in patterns between estimation methods suggests that use of a particular method can have predictable influence on results. Delta methods, and particularly the numerical variant, appear to approximate distributions reasonably well but may not capture the degree of asymmetry indicated by bootstrap and Bayes methods. Gavaris (1999) noted a similar pattern when comparing the analytical variant of the Delta method to bootstrap results. Consequently, inferences at low probability levels are likely to be inaccurate. Delta methods are however, simple to implement and the least compute intensive approach. The similarity in results between nonparametric percentile bootstrap and Bayes method could have been anticipated because non-informative priors were used and the maximum likelihood for lognormal index errors is equivalent to a nonparametric least squares solution on log transformed indices. This cannot be generalized however, and other distributions have been assumed for other stock assessments. Potential sensitivity to parametric assumptions about error distributions, as evidenced by differences between SEP/L and SEP/M results identifies a possible advantage of nonparametric approaches, an option not available in Bayesian methods. The close agreement between Bayes and *perc*NPB coupled with the difference between these and *bc*NPB suggests that estimation bias may displace distributions. Point estimates of bias using Box’s (1971) approximation and bootstrap compared favourably for these cases, suggesting that the bias was reasonably well determined. Handling this type of displacement for Bayes methods is unclear. Finally, though relative interest parameters do not offer any respite from differences between estimation methods, they appear to be more robust to structural model choice. Considering the impact of model choice on inferences, it may be well worth framing fisheries management advice in terms of relative measures when possible.

It is not clear how robust the parametric methods are to mispecification of error distributions. Nonparametric approaches are attractive because they relax the requirement to accurately specify the error distributions.

Conclusions

These results lead to the conclusion that choice of estimation method can have an appreciable impact on the perception of risks associated with the consequences of fisheries management decisions. Although further evaluation is required to understand the patterns of differences, some of the more regular features suggest the following preliminary interpretations. Delta methods did not capture the assymetry of distributions, thereby resulting in longer lower tails and smaller lower critical values for confidence intervals. This was unimportant for well estimated interest parameters, as in the plaice case. Bias adjustment is necessary to account for possible non-linearity induced displacement. Bias adjusted methods can shift distributions appreciably, though less so for relative quantities. The difficulty in addressing non-linearity induced bias with Bayesian methods is a concern.

Within a structural model, the range in percentiles for *SSB* (scaled to the average median) was fairly similar at the 25th, 50th and 75th percentiles, while it was typically, but not always, larger at the 5th and 95th percentiles. The scaled range at central probabilities was about 20% of the median except for plaice with the VPA/F structural model and sardine with the SEP/M structural model where it was less than 10%. The pattern was similar for range in *SSB* change but there was more diversity across structural models and cases. The range in *SSB* change at central probabilities varied from a few percent change to almost 100% change.

	Haddock		Plaice		Sardine		
	VPA/F	SEP/L	VPA/F	SEP/L	VPA/F	SEP/L	SEP/M
scaled range for *SSB*							
5	0.32	0.19	0.09	0.09	0.37	0.22	0.04
25	0.19	0.14	0.06	0.13	0.24	0.21	0.04
median	0.17	0.18	0.05	0.21	0.21	0.24	0.07
75	0.20	0.22	0.05	0.26	0.22	0.27	0.07
95	0.39	0.94	0.09	0.36	0.43	0.38	0.23
range for $(SSB_{1998} - SSB_{1992})/SSB_{1992}$							
5	0.89	1.17	0.08	0.13	0.37	0.20	0.01
25	0.52	1.02	0.05	0.11	0.24	0.19	0.03
median	0.46	0.93	0.04	0.11	0.21	0.16	0.01
75	0.46	0.81	0.04	0.14	0.22	0.15	0.03
95	1.02	1.36	0.07	0.20	0.43	0.19	0.10

Very broadly, these results suggest that the perceptions of probabilities and risks may be dependent on the chosen uncertainty assessment method by an amount of the order of 20% in the central part of the distributions, and that probabilities of the order of 5% and 95% are too dependent on methdology to be presented reliably.

Acknowledgements

The study has been carried out with financial support from the Commission of the European Communities, Agriculture and Fisheries (FAIR) specific RTD programme, CT98-4231, “Evaluation and comparison of methods for estimating uncertainty in harvesting fish from natural populations”. It does not necessarily reflect its views and in no way anticipates the Commission’s future policy in this area.

References

Box, M.J. 1971. Bias in nonlinear estimation. *Journal of the Royal Statistical Society, Series B* 33, 171-201.

Buckland, S.T., Burnham K.P. and Augustin, N.H. 1997. Model selection: an integral part of inference. *Biometrics* 53(2), 603-618.

Darby, C.D. and Flatman, S. 1994. Virtual Population Analysis: version 3.1 (Windows/Dos) user guide. *Information Technology Series, MAFF Directorate of Fisheries Research, Lowestoft* No. 1, 85pp.

Deriso, R.B., Quinn, T.J. II, and Neil, P.R. 1985. Catch-age analysis with auxiliary information. *Canadian Journal of Fisheries and Aquatic Science* 42, 815-824.

Efron, B. 1979. Bootstrap methods: another look at the jackknife. *Annals of Statistics* 7, 1-26.

Efron, B. 1982. *The jacknife, the bootstrap and other resampling plans (*CBMS Monograph No. 38). Society for Industrial and Applied Mathematics, Philadelphia.

Efron, B. 1998. R.A. Fisher in the 21st. century. *Statistical Science* 13(2), 95-122

Efron, B. and Tibshirani, R.J. 1993. *An introduction to the bootstrap*. Chapman and Hall, New York.

Fournier, D. and Archibald, C.P. 1982. A general theory for analyzing catch at age data. *Canadian Jounal of Fisheries and Aquatic Science* 39, 1195-1207.

Gavaris, S. 1988. An adaptive framework for the estimation of population size. *Canadian Atlantic Fisheries Scientific Advisory Committee Research Document* No. 88/29, 12 pp.

Gavaris, S. 1993. Analytical estimates of reliability for the projected yield from commercial fisheries. In: *Risk Evaluation and Biological Reference Points for Fisheries Management (*Canadian Special Publications in Fisheries and Aquatic Science, Vol.. 120) (eds S.J. Smith, J.J., Hunt, and D. Rivard) pp. 185-191.

Gavaris, S. 1999. Dealing with bias in estimating uncertainty and risk. In: *Providing Scientific Advice to Implement the Precautionary Approach Under the Magnuson-Stevens Fishery Conservation and Management Act. NOAA Technical Memorandum NMFS-F/SPO-40.* (ed V.R. Restrepo), US Department of Commerce, Washington, pp. 46-50.

Gavaris, S. and Van Eeckhaute, L 1998. Assessment of haddock on eastern Georges Bank. *Department of Fisheries and Oceans, Canadian Stock Assessment Secretariat Research Document* No. 98/66: 75pp.

Gilks, W. R., Richardson, S. and Spiegelhalter, D. J. 1996. Markov Chaine Monte Carlo in Practise. Chapmann & Hall, London.

Ianelli, J.N. and Fournier, D.A. 1998. Alternative age-structured analyses of the NRC simulated stock assessment data. In: *Analyses of Simulated Data Sets in Support of the NRC Study on Stock Assessment Methods*. *NOAA Technical Memorandum NMFS-F/SPO-30* (ed V.R. Restrepo), US Department of Commerce, Washington, pp. 81-96.

ICES. 1998. Report of the Working Group on the assessment of demersal stocks in the North Sea and Skagerrak. ICES C.M. 1998/Assess:7

ICES. 1999. Report of the Working Group on the assessment of mackerel, horse mackerel, sardine and anchovy. ICES C.M. 1999/Assess:6

Kennedy, W.J., Jr. and J.E. Gentle. 1980. Statistical computing. Marcel Dekker. New York. 591 pp.

McAllister, M.K., E.K. Pikitch, A.E. Punt, R. Hilborn. 1994. A Bayesian approach to stock assessment and harvest decisions using the samplin/importance resampling algorithm. *Canadian Jounal of Fisheries and Aquatic Science* 51: 2673-2687.

Mohn, R.K. 1993. Bootstrap estimates of ADAPT parameters, their projection in risk analysis and their retrospective patterns. In: *Risk Evaluation and Biological Reference Points for Fisheries Management (*Canadian Special Publications in Fisheries and Aquatic Science, Vol.. 120) (eds S.J. Smith, J.J., Hunt, and D. Rivard). pp. 173-187.

Patterson, K.R. 1999. Evaluating uncertainty in harvest control law catches using Bayesian Markov chain Monte Carlo virtual population analysis with adaptive rejection sampling and including structural uncertainty. *Canadian Jounal of Fisheries and Aquatic Science* 56: 208-221.

Patterson, K.R. and Melvin, G.D. 1996. Integrated catch at age analysis, version 1.2. *Scottish Fisheries Research Report* 58, 60 pp.

Patterson, K.R., R.M. Cook, C.D. Darby, S. Gavaris, L. Kell, P. Lewy, B. Mesnil, A.E. Punt, V. R. Restrepo, D.W. Skagen, and G. Stefansson. 1999. A review of some methods for estimating uncertainty in fisheries. Fisheries Research Services Report No. 7/99. Marine Laboratory, Aberdeen, Scotland.

Pope, J.G. 1972. An investigation of the accuracy of virtual population analysis using cohort analysis. *ICNAF Research Bulletin* 9: 65-74.

Ratkowsky, D.A. 1983. *Nonlinear regression modelling*. Marcel Dekker, New York.

Rubin, D. B. 1987. Comment: The calculation of posterior distributions by data augmentation. J. Am. Statist. Assoc. 82: 543-546.

Schweder, T. and N.L. Hjort 1999. *Frequentist analogues of priors and posteriors*. (Statistical Research Report No. 8), University of Oslo, Oslo.

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

Sinclair, A. and S. Gavaris 1996. Some examples of probabilistic catch projections using ADAPT output. DFO Atlantic Fisheries Research Documents 96/51, 12pp.

Smith, S.J. and Gavaris, S. 1993. Evaluating the accuracy of projected catch estimates from sequential population analysis and trawl survey abundance estimates. In: *Risk Evaluation and Biological Reference Points for Fisheries Management (*Canadian Special Publications in Fisheries and Aquatic Science, Vol.. 120) (eds S.J. Smith, J.J., Hunt, and D. Rivard) pp. 163-172.

Wade, P.R. 1999. A comparison of statistical methods for fitting population models to data. In Mcdonald, L. et al (eds). Marine mammal survey and assessment methods. Balkema

Walters, C.[J.] and Ludwig, D. (1994) Calculation of Bayes posterior probability distributions for key population parameters. *Can. J. Fish. Aquat. Sci.* **51**, 713-722.

Table 1. Summary of combinations of estimation methods with structural models that were analyzed for each case study.

Haddock						
	*an-shift*Delta	*num*Delta	*perc*PB	*perc*NPB	*bc*NPB	Bayes
VPA/F	X		X	X	X	X
SEP/L		X		X	X	X

Plaice						
	*an-shift*Delta	*num*Delta	*perc* PB	*perc* NPB	*bc*NPB	Bayes
VPA/F	X	X	X	X	X	X
SEP/L		X		X	X	X

Sardine						
	*an-shift*Delta	*num*Delta	*perc* PB	*perc* NPB	*bc*NPB	Bayes
VPA/F	X			X	X	X
SEP/L		X		X	X	X
SEP/M				X	X	X

*an-shift*Delta : bias adjusted analytical delta
*num*Delta : numerical delta
*perc*PB : parametric bootstrap
*perc*NPB : nonparametric bootstrap
*bc*NPB : bias adjusted nonparametric bootstrap
Bayes : Bayesian

Table 2. Priors used in the two Bayes variants for analysis of the haddock case using the VPA/F structural model. The notation *U(a,b)* denotes the uniform distribution on the interval from a to b.

Bayes1		***Bayes2***	
Parameter	***Prior distribution***	***Parameter***	***Prior distribution***
$\ln N_{1986,a}$: $a = 1,2,..,8$	*U(-∞, ∞)*	$N_{a,86}$: $a = 2,..,8$	*U(0,10^7)*
$\ln N_{y,1}$: $y = 1987,88,..,98$	*U(-∞, ∞)*	N_{1y} : $y = 87,...,98$	*U(0,10^7)*
$\ln F_y$: $y = 1986,87,..,97$	*U(-∞, ∞)*	F_y : y=86,…,97	*U(0,10)*
$\ln S_a$: $a = 1,2,..,7$	*U(-∞, ∞)*	s_a : a=2,…,7	*U(0,2)*
$\ln q_a$: $a = 1,2..,8$	*U(-∞, ∞)*	$q_{,a}$: a=2,…,8	*U(0,2)*
σ_1^2 *(index)*	$\propto \frac{1}{\sigma_1^2}$	σ_1^2 *(index)*	*U(0,2)*
σ_2^2 *(catch-at-age)*	$\propto \frac{1}{\sigma_2^2}$	σ_2^2 *(catch-at-age)*	*U(0,2)*

Table 3. Comparison of characteristics for the distribution of SSB in 1998 and for the distribution of change for SSB in 1998 relative to 1992 from combinations of structural model and uncertainty estimation method applied to the haddock case.

Structural Model	**VPA/F**					**SEP/L**				
Uncertainty Estimation	***an-shift* Delta**	***perc*PB**	***perc*NPB**	***bc*NPB**	**Bayes**	***num* Delta**	***perc*NPB**	***bc*NPB**	**Bayes1**	**Bayes2**
SSB_{1998}										
mean	34800	37874	42278	37627	41565	45819	50027	44077	50354	43449
median	34800	37757	41247	36712	40479	43877	43180	38905	46539	40880
Std. Dev.	9275	3896	9401	8463	9784	17271	25081	20145	17192	14295
5th percentile	19459	31640	28399	25304	27287	22521	27801	25418	30618	27739
25th percentile	28514	35087	35584	31772	34515	32792	35014	32620	38632	34898
75th percentile	41085	40367	47929	42632	46914	55308	54982	48935	58262	48930
95th percentile	50141	44534	59590	52868	59327	77689	104585	77630	82016	64660
scaled to median										
5th percentile	-15341	-6117	-12848	-11408	-13193	-21355	-15379	-13487	-15921	-13141
25th percentile	-6286	-2670	-5664	-4939	-5965	-11084	-8167	-6285	-7907	-5983
75th percentile	6286	2610	6681	5920	6434	11431	11802	10030	11723	8050
95th percentile	15341	6777	18343	16156	18848	33812	61405	38725	35477	23780
$(SSB_{1998}-SSB_{1992})/SSB_{1992}$										
mean	2.20	2.45	2.70	2.34	2.68	1.34	1.71	1.78	1.70	2.24
median	2.20	2.44	2.67	2.32	2.61	1.20	1.64	1.72	1.67	2.10
Std. Dev.	0.68	0.31	0.65	0.60	0.72	0.99	0.48	0.50	0.51	1.11
5th percentile	1.07	1.96	1.68	1.42	1.61	-0.06	1.00	1.06	0.95	1.11
25th percentile	1.74	2.23	2.26	1.91	2.16	0.64	1.36	1.43	1.35	1.66
75th percentile	2.66	2.65	3.08	2.72	3.11	1.90	1.99	2.09	2.00	2.71
95th percentile	3.33	2.97	3.85	3.41	3.99	3.18	2.58	2.66	2.57	3.93
scaled to median										
5th percentile	-1.13	-0.48	-0.99	-0.90	-0.99	-1.26	-0.64	-0.66	-0.71	-1.02
25th percentile	-0.46	-0.21	-0.40	-0.40	-0.44	-0.56	-0.28	-0.29	-0.32	-0.48
75th percentile	0.46	0.21	0.42	0.40	0.50	0.70	0.35	0.37	0.34	0.58
95th percentile	1.13	0.53	1.18	1.09	1.39	1.98	0.94	0.94	0.91	1.80

Table 4. Comparison of characteristics for the distribution of SSB in 1998 and for the distribution of change for SSB in 1998 relative to 1992 for combinations of somtructural model and uncertainty estimation method applied to the plaice case.

Structural Model	**VPA/F**						**SEP/L**				
Uncertainty Estimation	***an-shift* Delta**	***num Delta***	***perc*PB**	***perc*NPB**	***bc*NPB**	**Bayes**	***num* Delta**	***perc*NPB**	***bc*NPB**	**Bayes1**	**Bayes2**
SSB_{1998}											
mean	239236	253427	254402	252031	243955	250750	205465	246756	216878	245538	201831
median	239236	249954	251687	249377	241636	246545	204888	233010	211950	243999	198800
Std. Dev.	28667	29459	27301	30217	28184	31544	21347	102288	70293	41017	24069
5^{th} percentile	191819	211380	214502	209160	204357	206972	171419	183690	164510	184062	169400
25^{th} percentile	219808	231964	235391	231375	224854	228593	190194	211890	192490	214998	185600
75^{th} percentile	258663	271271	270918	268091	259502	268360	219378	259230	233330	270699	213400
95^{th} percentile	286652	304397	301788	305914	291844	307951	241703	317660	272830	318526	239300
scaled to median											
5^{th} percentile	-47416	-38574	-37185	-40217	-37279	-39573	-33469	-49320	-47440	-59937	-29400
25^{th} percentile	-19428	-17990	-16296	-18002	-16782	-17952	-14694	-21120	-19460	-29001	-13200
75^{th} percentile	19428	21317	19231	18714	17866	21815	14490	26220	21380	26700	14600
95^{th} percentile	47416	54443	50101	56537	50208	61406	36815	84650	60880	74527	40500
$(SSB_{1998}-SSB_{1992})/SSB_{1992}$											
mean	-0.21	-0.16	-0.15	-0.16	-0.19	-0.17	-0.38	-0.26	-0.28	-0.26	-0.38
median	-0.21	-0.18	-0.16	-0.17	-0.20	-0.18	-0.38	-0.28	-0.29	-0.27	-0.38
Std. Dev.	0.09	0.10	0.09	0.10	0.09	0.10	0.09	0.10	0.09	0.09	0.06
5^{th} percentile	-0.36	-0.30	-0.29	-0.30	-0.32	-0.31	-0.52	-0.39	-0.41	-0.41	-0.46
25^{th} percentile	-0.27	-0.23	-0.22	-0.23	-0.25	-0.24	-0.44	-0.33	-0.34	-0.33	-0.42
75^{th} percentile	-0.14	-0.11	-0.10	-0.11	-0.14	-0.11	-0.32	-0.22	-0.23	-0.21	-0.35
95^{th} percentile	-0.05	0.00	0.00	0.02	-0.03	0.02	-0.22	-0.09	-0.11	-0.10	-0.29
scaled to median											
5^{th} percentile	-0.16	-0.13	-0.12	-0.13	-0.12	-0.13	-0.15	-0.11	-0.12	-0.14	-0.08
25^{th} percentile	-0.06	-0.06	-0.05	-0.06	-0.06	-0.06	-0.06	-0.05	-0.05	-0.05	-0.04
75^{th} percentile	0.06	0.07	0.06	0.06	0.06	0.07	0.06	0.06	0.06	0.06	0.03
95^{th} percentile	0.16	0.18	0.16	0.20	0.17	0.21	0.16	0.19	0.18	0.18	0.09

Table 5. Comparison of characteristics for the distribution of SSB in 1998 and for the distribution of change for SSB in 1998 relative to 1992 from combinations of structural model and uncertainty estimation method applied to the sardine case.

Structural Model	**VPA/F**				**SEP/L**				**SEP/M**		
Uncertainty Estimation	***an-shift* Delta**	***perc*NPB**	***bc*NPB**	**Bayes**	***num* Delta**	***perc*NPB**	***bc*NPB**	**Bayes**	***perc*NPB**	***bc*NPB**	**Bayes**
SSB_{1998}											
mean	304	390	336	397	226	231	211	268	274	264	284
median	304	375	322	369	218	224	203	258	263	255	273
Std. Dev.	126	131	112	142	75	60	56	70	66	64	78
5^{th} percentile	96	218	191	222	122	149	139	172	181	173	183
25^{th} percentile	219	295	256	302	173	186	172	220	229	222	231
75^{th} percentile	390	452	392	464	267	266	244	305	307	297	316
95^{th} percentile	513	629	531	659	362	337	310	395	396	381	442
scaled to median											
5^{th} percentile	-208	-157	-131	-147	-96	-75	-65	-86	-83	-81	-90
25^{th} percentile	-85	-80	-66	-67	-45	-37	-32	-38	-35	-33	-42
75^{th} percentile	85	77	71	96	49	43	40	47	44	43	43
95^{th} percentile	208	254	209	290	144	113	107	137	132	127	169
$(SSB_{1998}$-$SSB_{1992})/SSB_{1992}$											
mean	-0.092	0.157	0.000	0.180	-0.396	-0.350	-0.373	-0.245	-0.131	-0.135	-0.134
median	-0.092	0.114	-0.038	0.098	-0.429	-0.375	-0.396	-0.266	-0.165	-0.169	-0.160
Std. Dev.	0.374	0.389	0.333	0.421	0.231	0.164	0.158	0.188	0.207	0.206	0.227
5^{th} percentile	-0.711	-0.357	-0.428	-0.342	-0.708	-0.572	-0.584	-0.509	-0.428	-0.429	-0.433
25^{th} percentile	-0.345	-0.124	-0.236	-0.101	-0.559	-0.465	-0.485	-0.374	-0.271	-0.275	-0.298
75^{th} percentile	0.162	0.344	0.170	0.381	-0.273	-0.258	-0.289	-0.135	-0.013	-0.016	-0.042
95^{th} percentile	0.527	0.865	0.583	0.959	0.029	-0.049	-0.087	0.107	0.238	0.228	0.327
scaled to median											
5^{th} percentile	-0.619	-0.471	-0.389	-0.440	-0.278	-0.197	-0.188	-0.243	-0.263	-0.260	-0.273
25^{th} percentile	-0.254	-0.237	-0.197	-0.199	-0.130	-0.091	-0.088	-0.108	-0.106	-0.106	-0.138
75^{th} percentile	0.254	0.231	0.209	0.283	0.157	0.117	0.108	0.131	0.152	0.153	0.118
95^{th} percentile	0.619	0.752	0.621	0.861	0.458	0.326	0.310	0.373	0.403	0.397	0.487

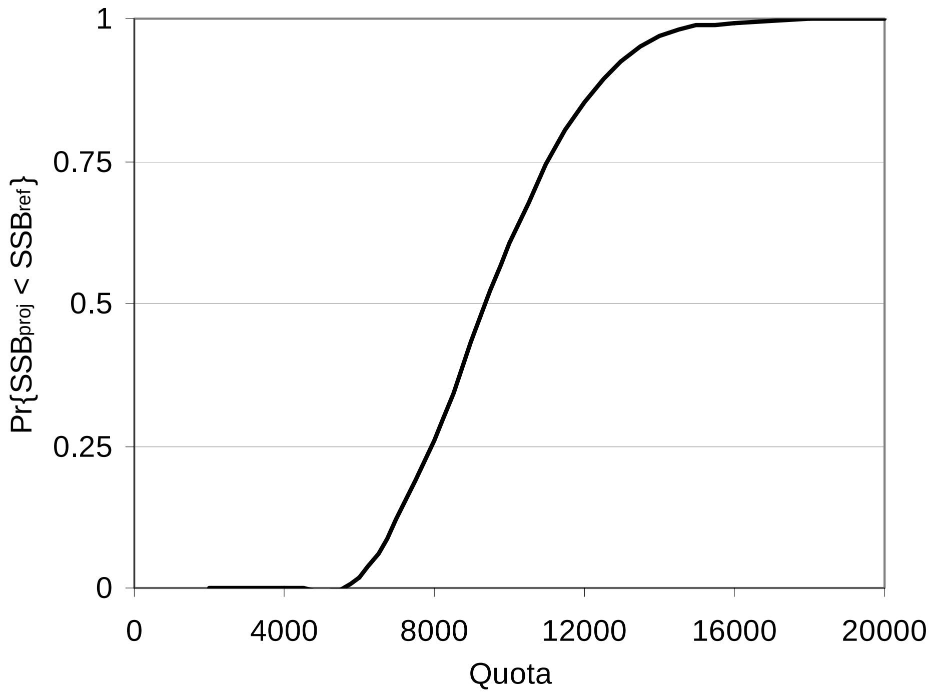

Figure 1. Characterization of the risk that projected spawning stock biomass will be less than its associated reference level for alternative catch quotas.

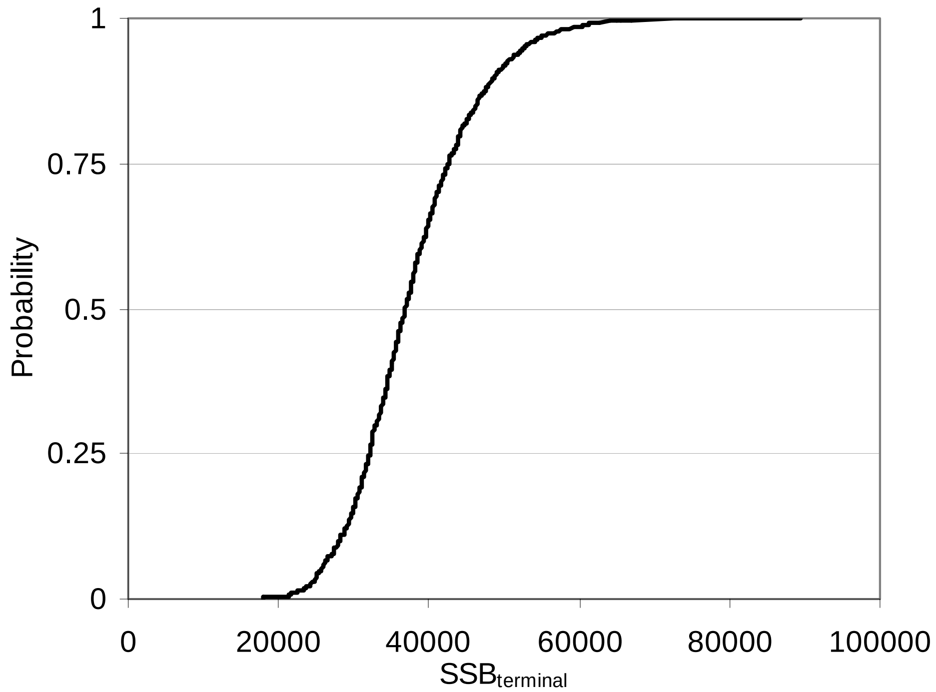

Figure 2. Confidence distribution (frequentist) or posterior distribution (Bayesian) for an interest parameter, in this example, spawning stock biomass in the terminal year.

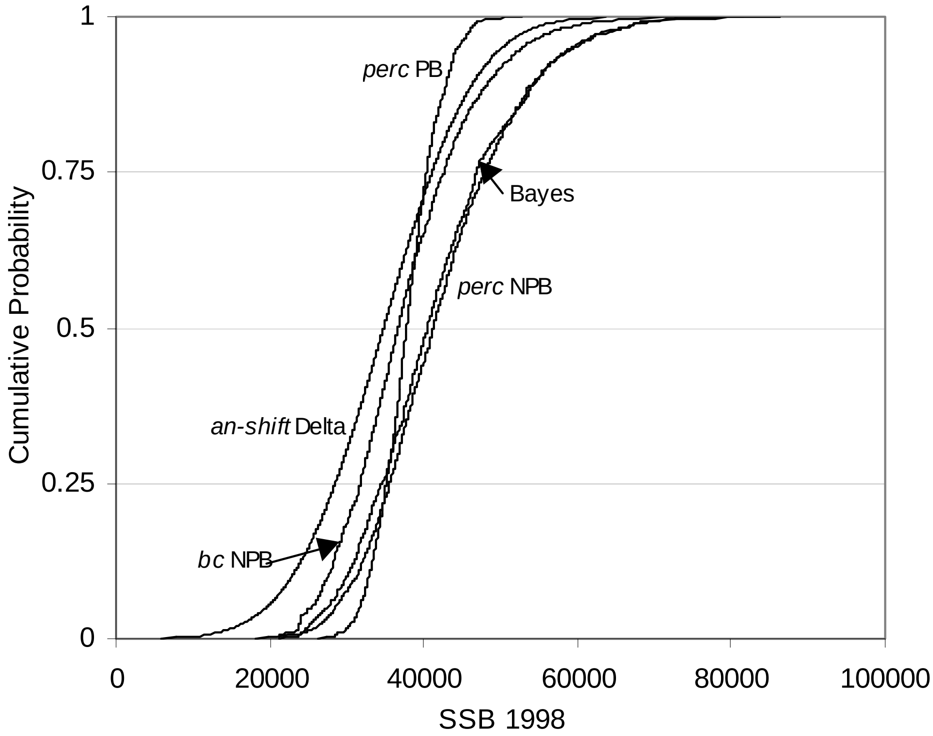

Figure 3. Distributions for haddock spawning stock biomass in 1998, calculated using a VPA/F structural model in combination with various methods of estimating uncertainty.

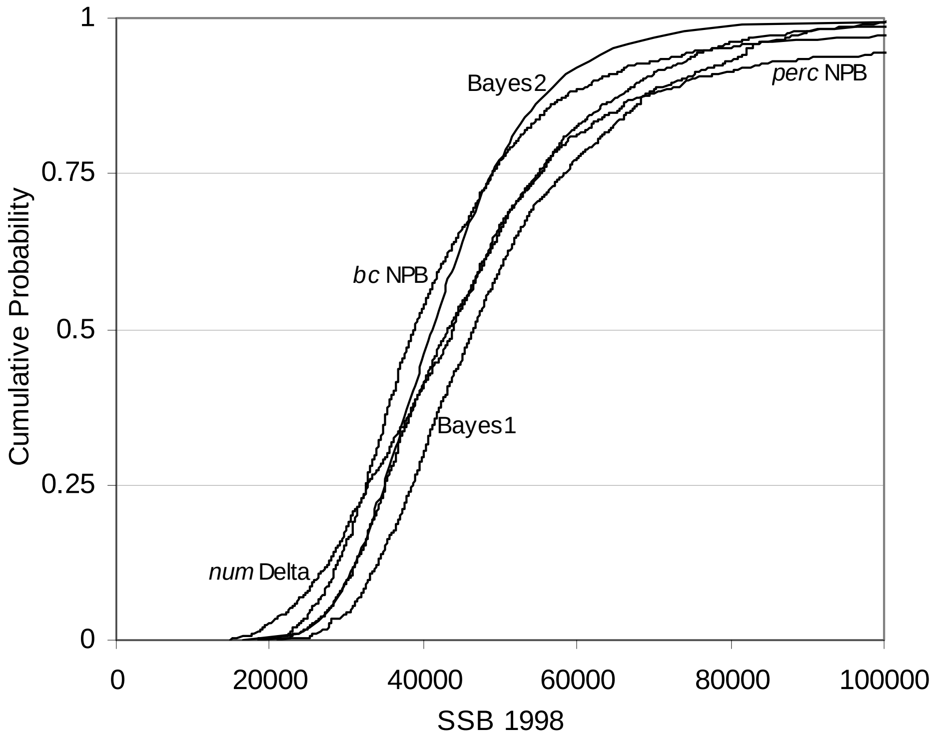

Figure 4. Distributions for haddock spawning stock biomass in 1998, calculated using a SEP/L structural model in combination with various methods of estimating uncertainty.

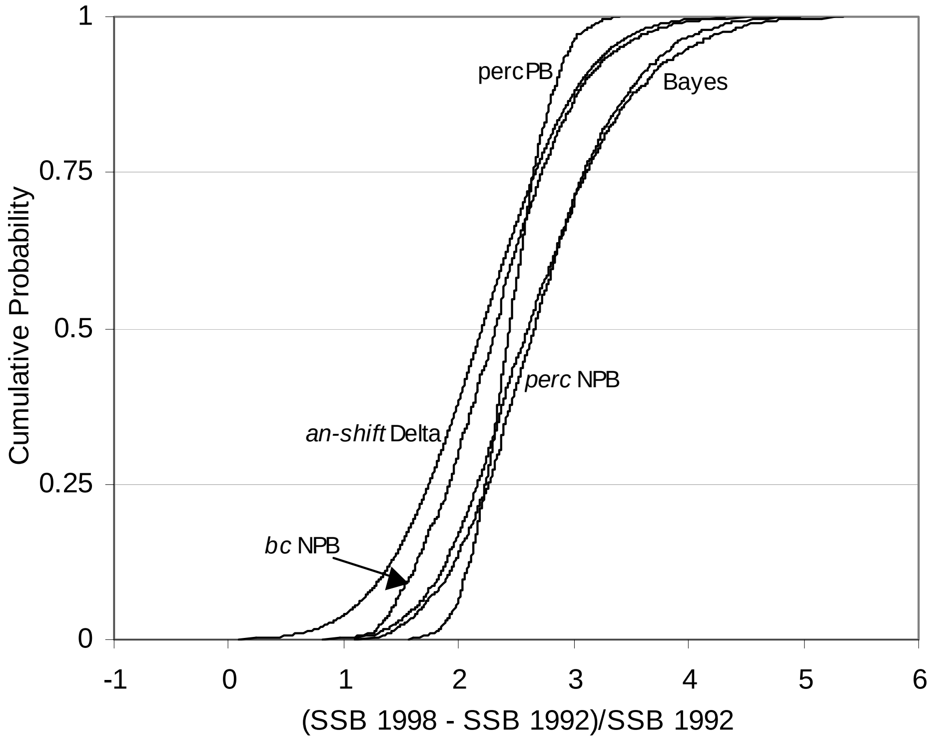

Figure 5. Distributions for change in haddock spawning stock biomass in 1998 relative to 1992, calculated using a VPA/F structural model in combination with various methods of estimating uncertainty.

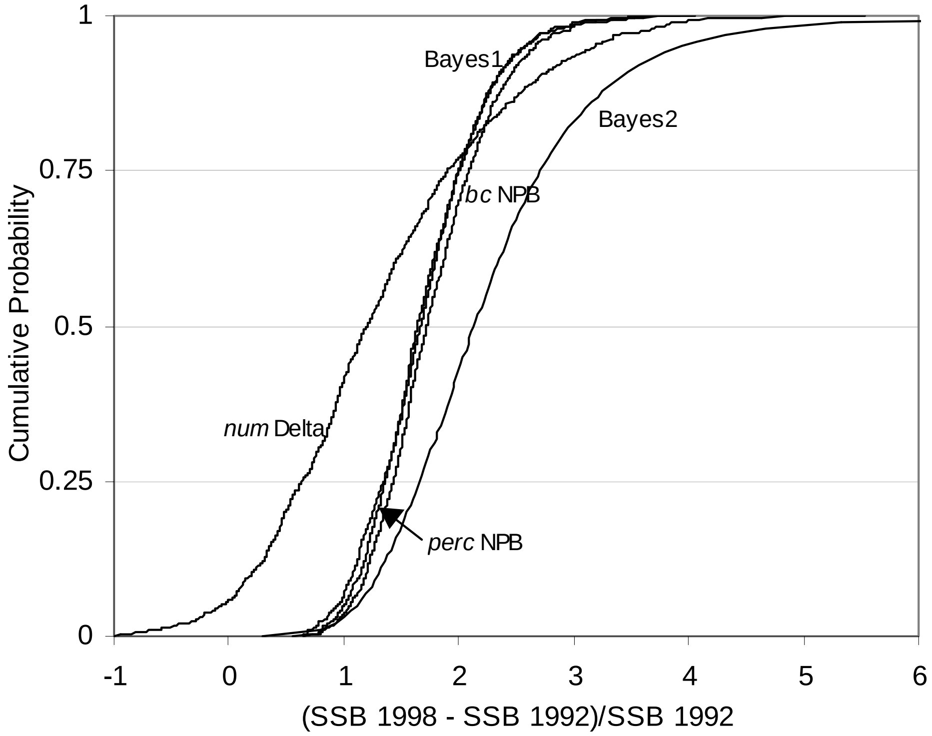

Figure 6. Distributions for change in haddock spawning stock biomass in 1998 relative to 1992, calculated using a SEP/L structural model in combination with various methods of estimating uncertainty.

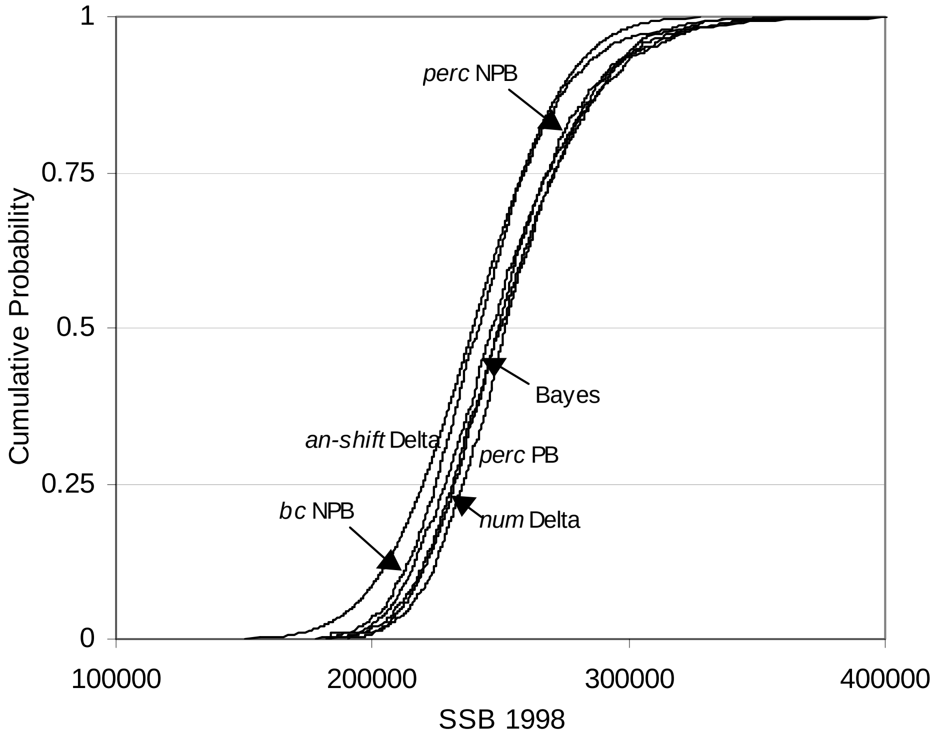

Figure 7. Distributions for plaice spawning stock biomass in 1998, calculated using a VPA/F structural model in combination with various methods of estimating uncertainty.

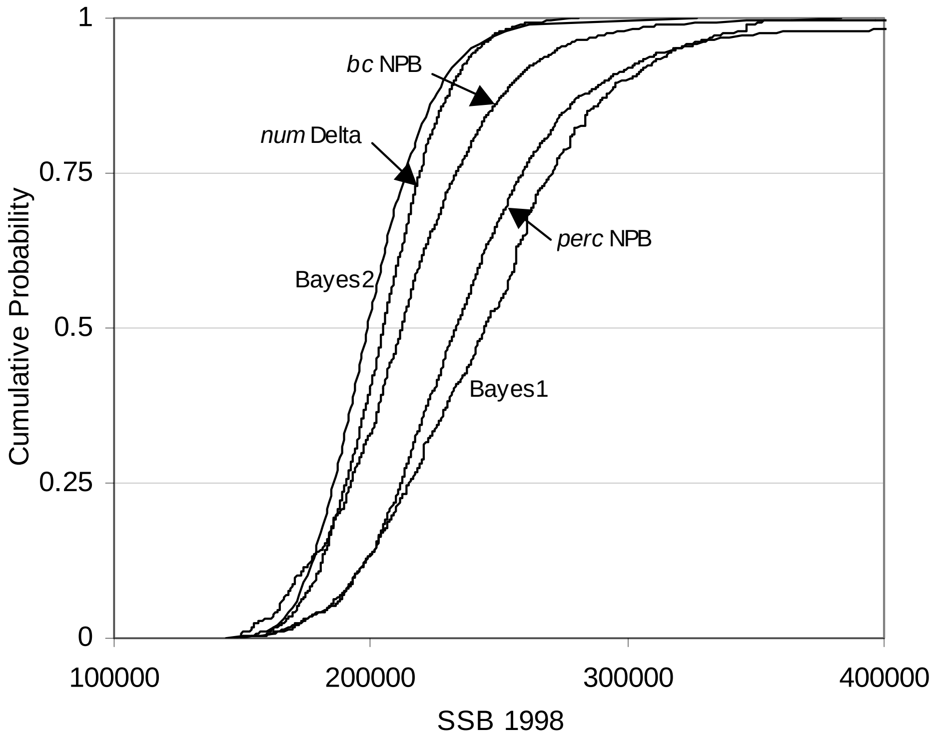

Figure 8. Distributions for plaice spawning stock biomass in 1998, calculated using a SEP/L structural model in combination with various methods of estimating uncertainty.

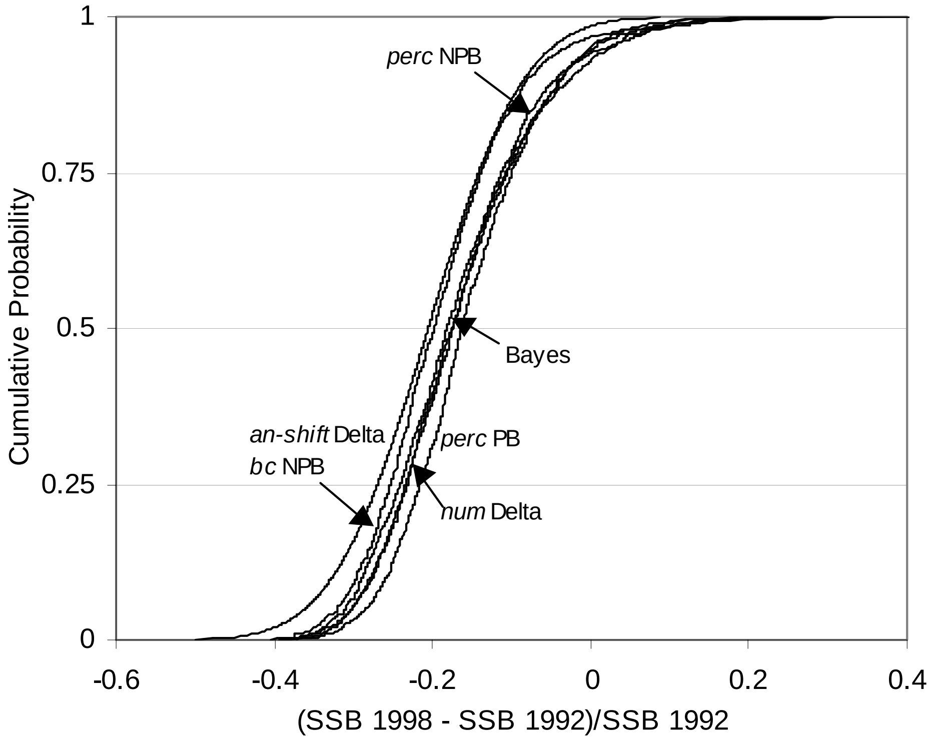

Figure 9. Distributions for change in plaice spawning stock biomass in 1998 relative to 1992, calculated using a VPA/F structural model in combination with various methods of estimating uncertainty.

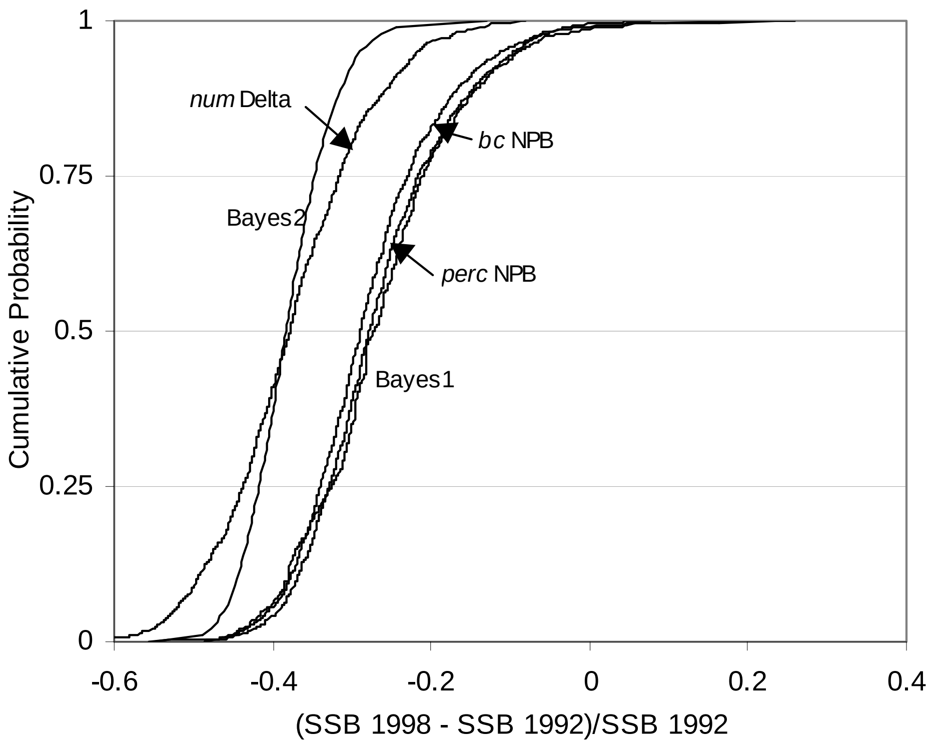

Figure 10. Distributions for change in plaice spawning stock biomass in 1998 relative to 1992, calculated using a SEP/L structural model in combination with various methods of estimating uncertainty.

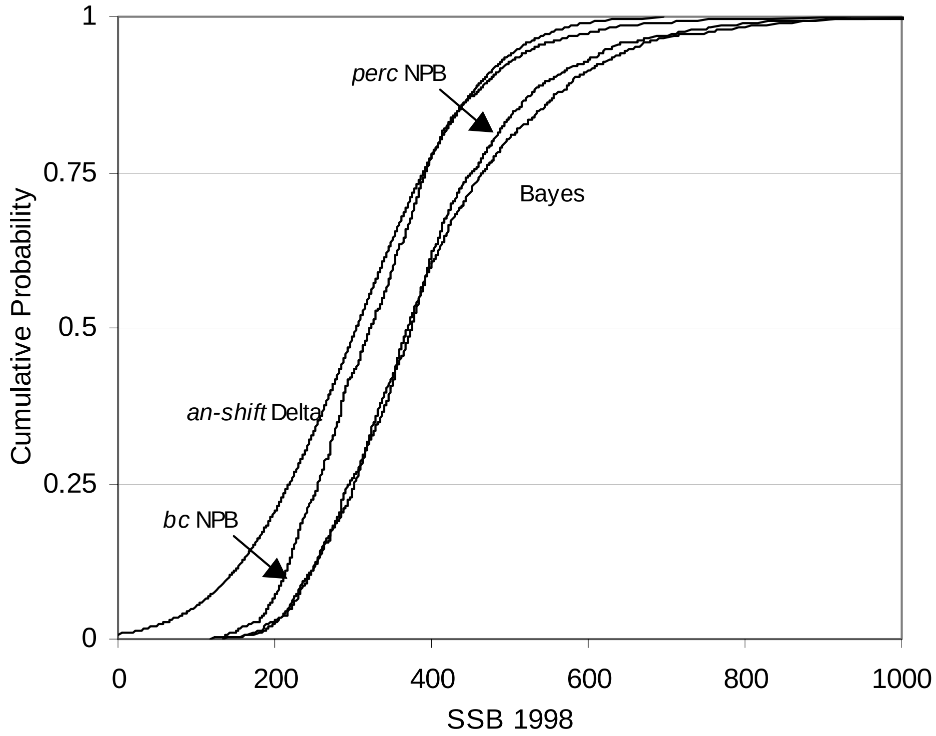

Figure 11. Distributions for sardine spawning stock biomass in 1998, calculated using a VPA/F structural model in combination with various methods of estimating uncertainty.

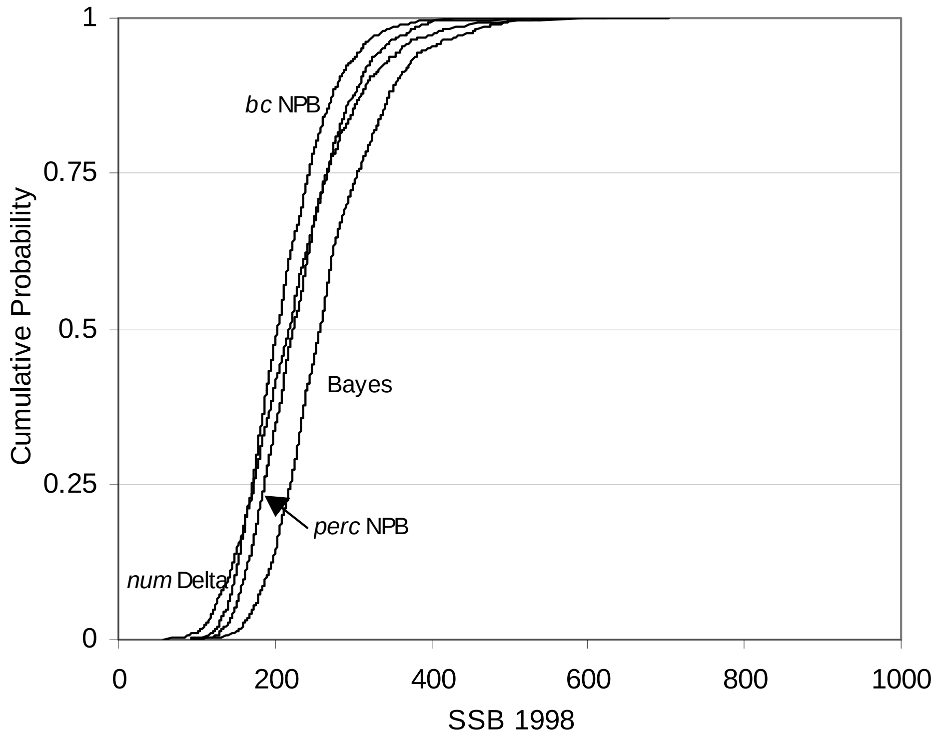

Figure 12. Distributions for sardine spawning stock biomass in 1998, calculated using a SEP/L structural model in combination with various methods of estimating uncertainty.

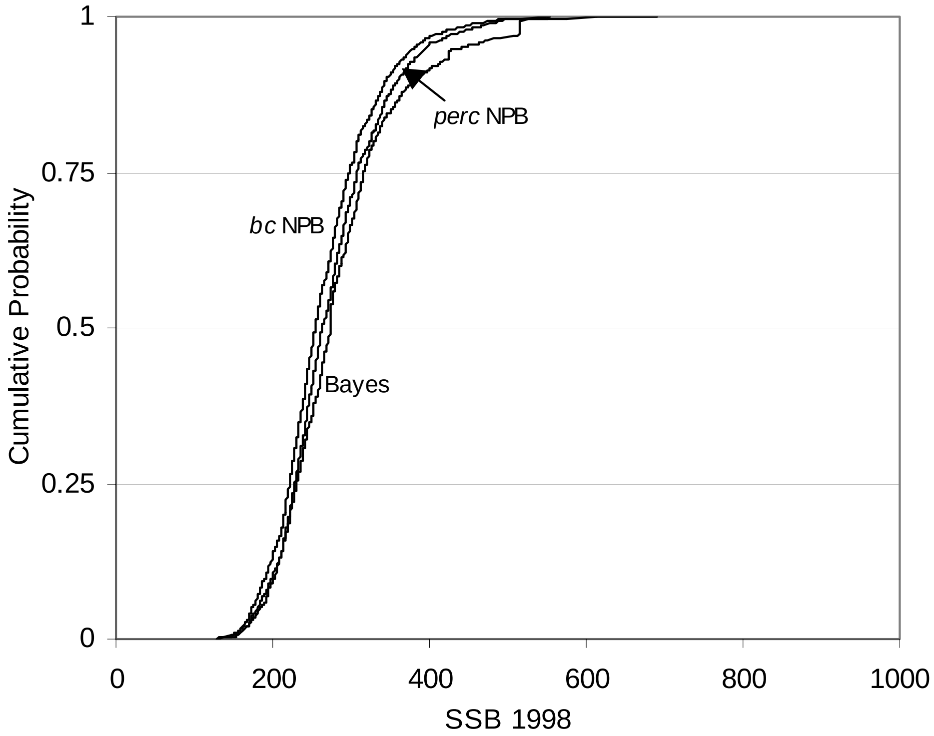

Figure 13. Distributions for sardine spawning stock biomass in 1998, calculated using a SEP/M structural model in combination with various methods of estimating uncertainty.

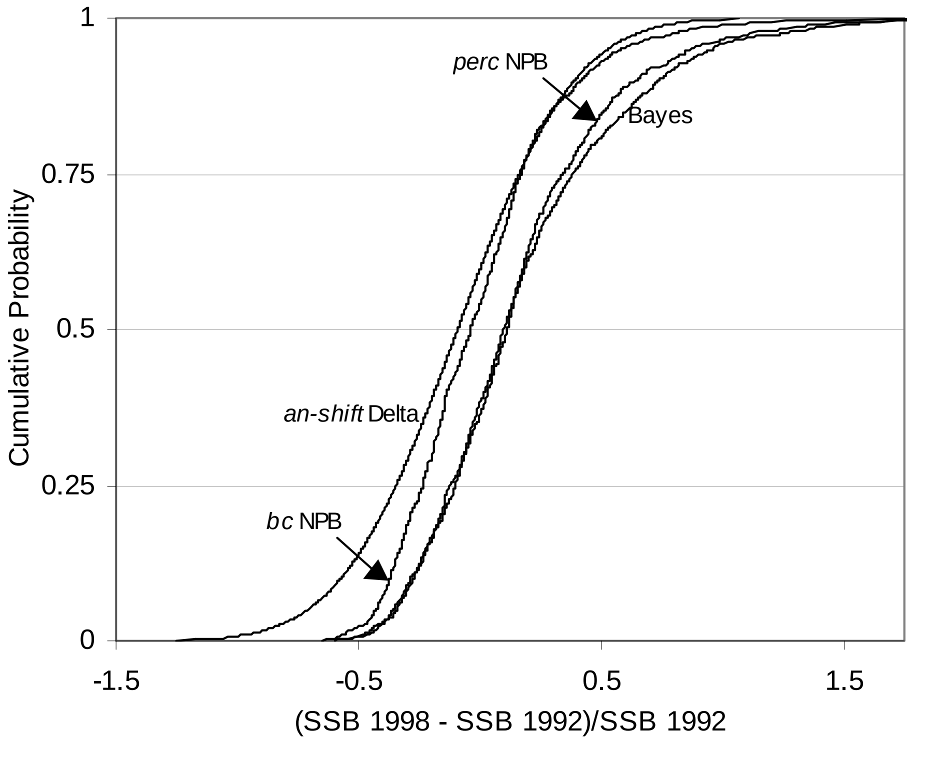

Figure 14. Distributions for change in sardine spawning stock biomass in 1998 relative to 1992, calculated using a VPA/F structural model in combination with various methods of estimating uncertainty.

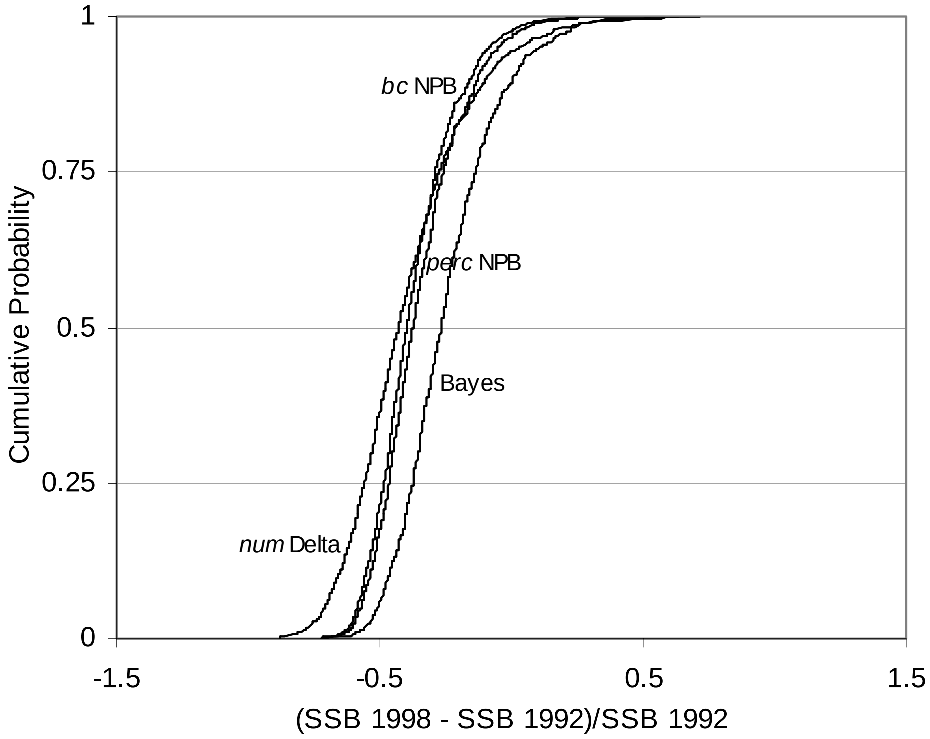

Figure 15. Distributions for change in sardine spawning stock biomass in 1998 relative to 1992, calculated using a SEP/L structural model in combination with various methods of estimating uncertainty.

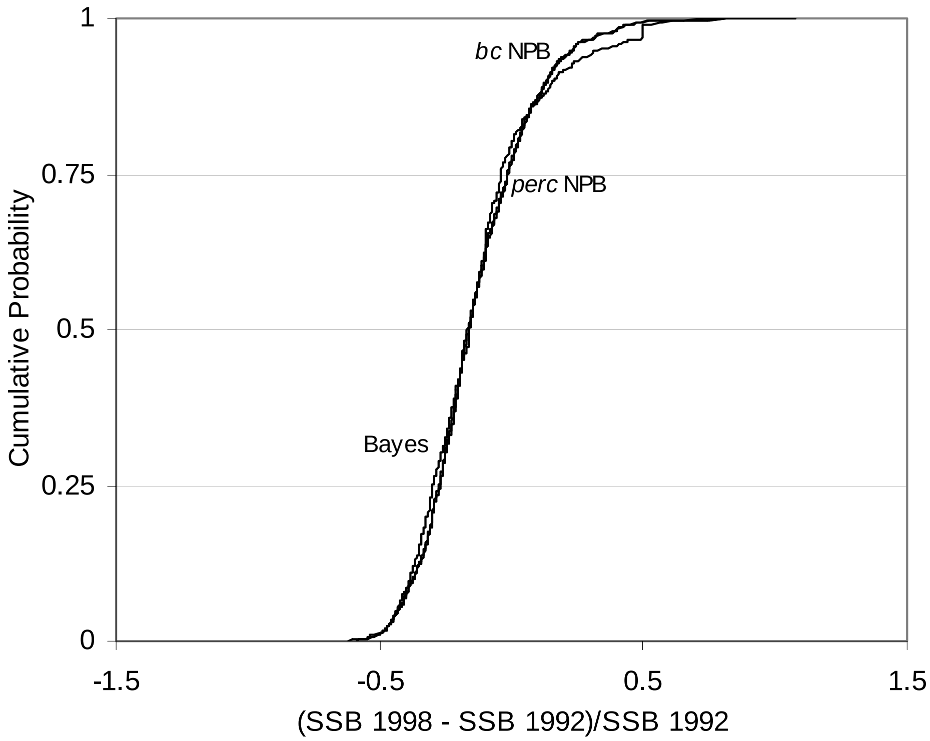

Figure 16. Distributions for change in sardine spawning stock biomass in 1998 relative to 1992, calculated using a SEP/M structural model in combination with various methods of estimating uncertainty.

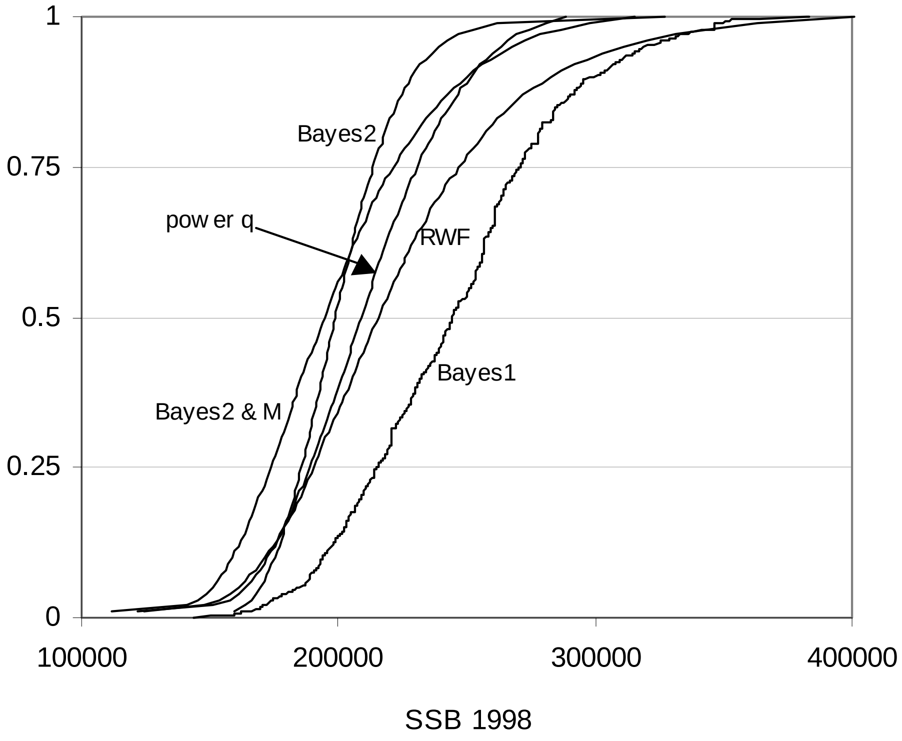

Figure 17. Distributions for plaice spawning stock biomass in 1998, calculated using a SEP/L structural model in combination with the Bayes method of estimating uncertainty. Details of the structural model and assumptions were altered.

Annex 1

A potential problem for generating model conditioned nonparametric bootstrap sample replicates arises when dealing with data that are assumed to be multinomially rather than normally / lognormally distributed. Consider the case in which catch-at-age data are included as one component of a likelihood function:

$$L = -\sum_y \sum_a C_y \, p_{y,a} \, \ell n \hat{p}_{y,a} \tag{A.1}$$

where C_y is the total catch in number for year y,

$p_{y,a}$ is the observed proportion of the catch of age a during year y, and

$\hat{p}_{y,a}$ is the model-estimate of the proportion of the catch of age a during year y.

This equation is not of the form of a sum of squared residuals, so the typical approach of scaling weighted residuals cannot be applied exactly. However, it is possible to define a standardised residual of the form:

$$r_{y,a} = \frac{p^{obs}_{y,a} - \hat{p}_{y,a}}{\sqrt{\hat{p}_{y,a}(1-\hat{p}_{y,a})}} \tag{A.2}$$

'Raw' pseudo catch proportions-at-age are generated using the equation:

$$p^{raw,B}_{y,a} = \hat{p}_{y,a} + r^B_{y,a}\sqrt{\hat{p}_{y,a}(1-\hat{p}_{y,a})} \tag{A.3}$$

where $r^B_{y,a}$ is a residual selected randomly from those defined in Equation (A.2). The 'final' pseudo catch proportions-at-age are then determined by rescaling the 'raw' catch proportions-at-age so that the sum over age (within each year) of the 'final' pseudo catch proportions-at-age is 1, i.e.:

$$p^B_{y,a} = p^{raw,B}_{y,a} / \sum_{a'} p^{raw,B}_{y,a'} \tag{A.4}$$