Final Aug 30/2000

Do Different Methods Provide Accurate Probability Statements in the Short Term?

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

Abstract.- The performance of uncertainty estimation procedures was evaluated with respect to accuracy. A confidence statement is said to be accurate if the confidence point achieves the desired probability coverage. A Monte Carlo experiment with 100 trials was conducted with a "true" population that experienced contrast between low and high fishing mortality. Observations for the last 25 years were drawn stochastically by adding measurement error. The assessment approaches were VPA-based (ADAPT and XSA with errors on the effort data). The "Delta Method", parametric bootstrap and non-parametric bootstrap (NPB), and a Bayesian approach were used to quantify coverage and assess the accuracy of confidence limits of estimated interest parameters ($F_{0.1}$, SSB and $TAC_{F0.1}$ in year 26) by comparing against the "true" values. Variations of the Delta Method and bootstrap were used to account for statistical estimation bias. The results indicated that accurate inference statements are possible with the different approaches and that bias correction can improve accuracy when it can be applied. The bias-corrected Delta-ADAPT and bias-corrected NPB-ADAPT applications performed best. Inference statements about $F_{0.1}$ were more accurate than those for SSB or TAC.

V. Restrepo: International Commission for the Conservation of Atlantic Tunas, Corazón de María 8-Sixth, 28002 Madrid, Spain [tel. +34 91 4165600, fax +34 91 4152612, e-mail: victor.restrepo@iccat.es]. K. Patrterson and R. Cook: FRS Marine Laboratory, Aberdeen, UK. C. Darby, L. Kell and C. O´Brien: CEFAS Laboratory, Lowestoft, UK. S. Gavaris: DFO Biological Station, St. Andrews, Canada. P. Lewy: DIFRES, Charlottenlund, Denmark. B. Mesnil: IFREMER, Nantes, France. A. Punt: CSIRO Marine Research, Hobart, Australia. D. Skagen: IMR, Bergen, Norway. G. Stefánsson: MRI and UI, Reykjavik, Iceland.

INTRODUCTION

Before evaluating the performance of any uncertainty estimation procedure in real situations, it is a good idea to evaluate it in controlled experiments, as are afforded by Monte Carlo simulation studies.

In this study we are interested in evaluating the accuracy of various estimation procedures given that the assessment method in which they are embedded make the same structural assumptions made in generating the simulated data. That is, we try to avoid confounding of these two potential sources of inaccuracy: the mismatch between assumed and real processes, and the intrinsic differences in uncertainty estimation procedures. Naturally, assumptions about the way the real world operates are a very important part of stock assessments and cannot be ignored in real applications, but our more narrowly-focused study should be a first step in this type of performance evaluation.

One way of categorizing age-structured assessment methods in vogue today is as either admitting or ignoring measurement errors in catch-at-age observations. It is well known that the analysis of age-structured information is over-parameterized and both of these types of approaches aim to reduce the number of parameters to be estimated by making particular assumptions. Methods that do not admit catch-at-age errors are sometimes called "VPA-based". By not admitting catch errors, they greatly reduce the number of parameters that require estimation. Separable methods allow for the statistical modeling of the admitted error in the catch-at-age observations, thus increasing the number of data being "fitted" by the model, but reduce the number of estimated parameters by assuming that fishing mortality can be split into two-way age-specific and year-specific components. The analyses conducted for this study are done with "VPA-based" methods only. Although we did not examine separable-type models, testing based on these could be conducted in a similar fashion.

METHODS

Monte Carlo Test Procedure

Our study examines the quality of inference statements judged with respect to *accuracy*. A confidence statement is said to be accurate if the confidence point achieves the desired probability coverage, $\Pr(\eta \leq \hat{\eta}[\alpha]) \approx \alpha$.

To compare the accuracy of alternative methods of making inference statements, a Monte Carlo experiment with 100 trials was conducted. The specification of the simulation structure for this experiment is described below. Each of the data sets was analyzed using an assessment procedure that made the same assumptions made in generating the data. The "Delta Method", parametric bootstrap and non-parametric bootstrap (NPB), and a Bayesian approach were used to quantify coverage and assess the accuracy of confidence limits of estimated interest parameters.

The results from the Monte Carlo experiment can be used to quantify coverage and assess the accuracy of confidence limits. For example, under the frequentist paradigm, using a correct model we would expect 10 of the 100 cumulative frequency distributions to contain the 'true' *SSB* between the 20th and 30th percentile confidence limits and similarly for each of the ten deciles from 0% to 100% (Figure 1). These same results can be plotted on a cumulative basis (Figure 1), in which case one would expect a straight line.

Using an incorrect model for inference statements would result in patterns that would deviate from the straight lines shown in Figure 1. Differences in accuracy could be attributed to either

differences in the location of the point estimate with respect to the true value, or differences in spread, or both. Figure 2 depicts eight possible diagnostic patterns, labelled A to H. Note that in practice it may be difficult to distinguish the cause of patterns C, E or G based on these plots alone, because they all look similar.

Dynamics of simulated population

The dynamics for the stock have fixed growth, maturity and natural mortality:

Length at age: $l_a = 100(1 - e^{-0.15a})$

Weight at age: $w_a = 0.00001 \, l_a^{\,3}$
(for the purpose of computing yield, mid-year weight computations are made, i.e. using [a+0.5]; for computing SSB, beginning of the year [a] is used)
M=0.2 for all ages, 1 to 15 (there is no plus group)
Proportion mature:

0,	a=1 to 4,
0.3,	a=5
0.5,	a=6
0.7,	a=7
0.9,	a=8
0.95,	a=9
1,	a=10 and older.

The only stochastic process was in recruitment, which was governed by a Beverton-Holt relationship with a steepness of 0.7 having lognormal errors (CV=0.6) and following an AR(1) process with ρ=0.5. The deterministic relationship was given by

$$R_{t+1} = \frac{SSB_t}{0.67945 + 5.6621 \times 10^{-5} \, SSB_t},$$

which results in a virgin biomass of 100,000, and a maximum recruitment of 15,769.

A single fleet with fixed selection pattern over time,
    S_a = 0.05, 0.1, 0.3, 0.7, 0.9, for a=1 to 5 and S_a=1.0 for older ages,
exploits the stock. Catchability was held constant and there were no stochastic components to the fleet dynamics.

With the above parameters, the following reference points are estimated:

F_{max}	0.335
$F_{0.1}$	0.193
F_{MSY}	0.166
F_{crash}	0.473.

Initially, the operating model was run until the population reached equilibrium at ½F_{MSY}. Then, over a 25-year period, the stock experienced a "2-way trip" to 2F_{MSY} and back to ½F_{MSY}. The purpose of this was to allow substantial contrast in the data.

Observations used for model fitting

For all 25 years, surveys (CPUE) indexing start of the year abundance for ages 1-15 were generated. These had lognormal errors with age-specific CVs ranging between 0.5 and 0.75 (higher CVs for young and old ages, lower CVs for intermediate ages). No survey data were produced for year 26.

For testing of the VPA-based models, the true catch at age (ages 1-15, without a plus group) was used.

Performance Statistics

Although the individual assessment realizations generated a large number of outputs, we limited the analysis of results to the following three interest statistics:

- SSB at start of year 26 (true = 31,801),
- $F_{0.1}$ (calculated using the estimated selectivity for year 25, true = 0.193), and
- TAC in year 26 corresponding to the estimated $F_{0.1}$ (true = 9,334).

The rationale for selecting these three was the following: SSB is one of the population variables that is monitored in most assessments, as it indexes the reproductive potential of the stock at any one time. $F_{0.1}$ is a biological reference point that is used as a target by various fishery management organizations. Lastly, the TAC corresponding to the target fishing mortality is the quantity that managers are ultimately interested in.

Computation of the TAC in year 26 required a projection of recruits, which was fixed at zero for simplicity.

Inference Methods

We examined frequentist and Bayesian inference methods. Frequentist confidence distributions of interest parameters are derived from sampling distributions of their estimators (Efron 1998, Schweder and Hjort 1999) and represent the probability that the true value of the interest parameter (assumed to be a single valued constant) would be contained within the specified limits under repeated sampling. Bayesian posterior distributions of interest parameters are derived from a synthesis of the likelihood for the observed data and specified prior distributions for model parameters (Punt and Hilborn 1997) and represent the probability distribution of the interest parameter (assumed to be a random variable). Though the frequentist confidence distribution and the Bayesian posterior distribution have different interpretations, both are used in a similar way to support fisheries management decisions. For the purposes of evaluating accuracy, analysts provided the percentile in which the true value of the statistic (SSB, $F_{0.1}$ or TAC) fell in each of the 100 trials.

Various techniques have been devised to obtain inferences under the frequentist or Bayesian paradigms. The frequentist methods examined included Delta and bootstrap variants while Bayesian methods were based on numerical integration using either Sampling Importance Resampling (Rubin 1987) or Markov Chain Monte Carlo algorithms (Gilks et al. 1996).

The Delta method is a technique for deriving approximate estimates of variance for complex model parameters and/or interest statistics. These estimates of variance may then be coupled

with some assumption about the sampling distribution of the model parameter or interest statistic to construct confidence distributions.

The bootstrap is a data based simulation technique that can be used to obtain confidence distributions of interest parameters. The idea is to substitute a simple data based estimate for the sampling distribution of a parameter. Parametric variants of the bootstrap assume a parametric form of the distribution but obtain estimates of the defining parameters from observed data. Non-parametric variants of the bootstrap use the observed data, or residuals about the model fit, directly to define the distribution completely.

Sampling Importance Resampling and Markov Chain Monte Carlo are techniques for integrating over the posterior distribution of model parameters given the observed data. 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.

Fisheries assessment models involve relationships that are not linear in the model parameters and the interest parameters. Estimation for such models will result in confidence distributions or posterior distributions that are displaced. The frequentist notion associated with this characteristic is bias. Some adjustment for statistical estimation bias was attempted with several of the frequentist methods, and in those cases the results are presented separately. Bayesian analogues for adjustment of a displacement are not available.

VPA-type Assessment Models

Although VPA based methods succeed in reducing the number of parameters to be estimated by assuming that the error in the catch at age is negligible, experience has shown that further assumptions are often required to obtain reliable estimates of stock status. Two variants of VPA type methods are common. One imposes additional conditioning by constraining the oldest age fishing mortality to be equal to some function of fishing mortality for younger ages in the same year. The other imposes additional conditioning by constraining the index catchabilities to be equal over specified older ages. As indicated earlier, the intention was to apply assessment methods which made correct structural assumptions. For this simulation study, assessment methods were tailored to make the correct assumption that expected fishing mortality for age 15 was equal to fishing mortality over younger ages, i.e. ages 6 - 14, in the same year. All index catchabilities for ages 1 - 15 were estimated independently without constraints.

Additional structural assumptions required for the assessment model were made to be consistent with the operational model. The annual natural mortality rate, M, is assumed constant and equal to 0.2 over all ages and years. The errors in the abundance indices, $I_{a,y}$, are assumed independent and identically distributed with age specific variance across all years, σ_a^2, after taking natural logarithms of the values. The methods were tasked with estimating age-specific CVs (i.e. make the same structural assumption made during data generation, but were not given the true values to use as input).

The F-constrained VPA assessments were carried out using either the ADAPT adaptive framework (Gavaris 1988) with various software implementations or the XSA extended survivors iterative algorithm (Shepherd 1999) as implemented in the Lowestoft assessment suite (Darby and Flatman 1994). It should be noted that there are differences in the implementation of various algorithms by different scientists (for instance in stopping criteria) which can lead to differences in the results that could complicate interpretation of results. We carried out simple

comparisons of the individual iteration results obtained with the various ADAPT and XSA implementations and found that, while differences existed, they were not substantial.

ADAPT specification

Model parameters:

$\theta_{a,26} = \ell n \, N_{a,26} = \ell n$ population abundance, ages $a = 2 - 15$ at the beginning of year 26,

$\kappa_a = \ell n \, q_a = \ell n$ calibration constant, for ages $a = 1 - 15$.

A weighted least squares solution for the parameters was obtained by minimizing the sum of squared differences between the natural logarithm of observed abundance indices and the natural logarithm of population abundance adjusted for catchability by the calibration constants:

$$\Psi_{a,y}\left(\hat{\theta}, \hat{\kappa}\right) = \sum_{a,y} \frac{1}{\hat{\sigma}_a^2} \left(\ell n \, I_{a,y} - \left(\hat{\kappa}_a + \ell n \, N_{a,y}\left(\hat{\theta}\right)\right)\right)^2$$

or equivalently, for the Bayesian methods, using the likelihood function:

$$L(D \mid \theta, \kappa) \propto \prod_{a,y} \tfrac{1}{\hat{\sigma}_a} \exp(-\tfrac{1}{2(\hat{\sigma}_a)^2}[\ell n \, I_{a,y} - \ell n\left(q_a N_{a,y}\right)]^2) \,.$$

At the beginning of year 26, the population abundance was obtained directly from the parameter estimates, $N_{a,26} = e^{\hat{\theta}_{a,26}}$. For all other times, the population abundance was computed using the VPA algorithm, which incorporates the common exponential decay model

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

The annual fishing mortality rate, $F_{a,y}$, was obtained by solving the catch equation,

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

The fishing mortality rate for age 15 was assumed equal to the average for ages 6 - 14 during that time interval

$$F_{15,y} = \sum_{a=6}^{14} F_{a,y} \Big/ 9 \,.$$

The interest statistics, SSB_{26}, $F_{0.1}$ and TAC_{26} were subsequently calculated.

$$SSB_{26} = \sum_a m_a w_a N_{a,26}$$

where m is proportion mature at age and w is weight at age.

$F_{0.1}$ is the fishing mortality corresponding to one tenth the slope at the origin of the yield per recruit as a function of fishing mortality. Yield per recruit is

$$YPR = \sum_a w_{a+0.5} N_a \, FPR_a \left(1 - e^{-(FPR_a + M)}\right) / FPR_a + M$$

where N_1 is set to 1 and the partial recruitment to the fishery is derived from the estimated fishing mortality at age for year 25

$$PR_a = F_{25} \Big/ \frac{1}{10} \sum_{a=6}^{15} F_a$$

The TAC corresponding to $F_{0.1}$ is

$$TAC_{26} = \sum_a w_{a+0.5} N_{a,26} \, F_{0.1} PR_a \left(1 - e^{-(F_{0.1} PR_a + M)}\right) / F_{0.1} PR_a + M$$

Delta- ADAPT

This implementation of the Delta method derives approximate estimates of variance for the interest parameters and couples that with an assumption about the parametric form of their sampling distribution to derive confidence distributions. The covariance matrix of the model parameter set, $\{\theta, \kappa\}$ were estimated using the common linear approximation (Kennedy and Gentle 1980 p.476)

$$\text{cov}(\hat{\theta}, \hat{\kappa}) = \hat{\sigma}^2 \left[J^T(\hat{\theta}, \hat{\kappa}) J(\hat{\theta}, \hat{\kappa}) \right]^{-1}$$

where $\hat{\sigma}^2$ is the mean square residual and $J(\hat{\theta}, \hat{\kappa})$ is the Jacobian matrix of the vector of residuals. The variance of an interest parameter, $\hat{\eta} = g(\hat{\theta}, \hat{\kappa})$ where g is the transformation function, were estimated using the Delta approximation (Ratkowsky 1983):

$$Var(\hat{\eta}) = tr\left[GG^T \text{cov}(\hat{\theta}, \hat{\kappa}) \right]$$

where G is the vector of first derivatives of g with respect to parameters. Assuming a Gaussian distribution, confidence distributions of the interest parameters were approximated as

$$N \sim \left(\hat{\eta}, \sqrt{Var(\hat{\eta})}\right).$$

As indicated above, estimation bias is expected. A bias adjusted Delta confidence distribution was constructed by shifting the Delta confidence distribution to account for the magnitude of the estimated bias and ignoring any increase in variance associated with the bias estimate. The bias of the model parameters was estimated using Box's (1971) approximation, which assumes that the errors are normally distributed:

$$Bias(\hat{\theta}, \hat{\kappa}) = \frac{-\hat{\sigma}^2}{2} \left(\sum_{i'} J_{i'}(\hat{\theta}, \hat{\kappa}) J_{i'}^T(\hat{\theta}, \hat{\kappa}) \right)^{-1} \left(\sum_i J_i(\hat{\theta}, \hat{\kappa}) \right) tr\left[\left(\sum_{i'} J_{i'}(\hat{\theta}, \hat{\kappa}) J_{i'}^T(\hat{\theta}, \hat{\kappa}) \right)^{-1} H_i(\hat{\theta}, \hat{\kappa}) \right]$$

where $J_i(\hat{\theta}, \hat{\kappa}) = J_{i'}(\hat{\theta}, \hat{\kappa})$ are vectors of the first derivatives for each residual and $H_i(\hat{\theta}, \hat{\kappa})$ are the Hessian matrices of second derivatives for each residual. The bias of interest parameters is then derived using the method described in Ratkowsky (1983):

$$Bias(\hat{\eta}) = G^T Bias(\hat{\theta}, \hat{\kappa}) + tr\left[W \, \text{cov}(\hat{\theta}, \hat{\kappa}) \right] / 2$$

where W is the matrix of second derivatives of g with respect to parameters. Again assuming a Gaussian distribution, confidence distributions of the interest parameters were approximated as $N \sim \left(\hat{\eta} - Bias(\hat{\eta}), \sqrt{Var(\hat{\eta})} \right)$.

PB-ADAPT

This method is explained in Restrepo et al. (1992) and Mesnil (1995). The method requires externally provided CVs in order to generate replicates of the dataset. Here, the same CVs as those used in generating the artificial data sets were used.

NPB-ADAPT

Efron (1979) introduced the non-parametric bootstrap percentile method as a simple automatic data based technique to calculate confidence distributions for complex statistics. Let η represent the interest parameter, (with estimate $\hat{\eta}$ corresponding to the least-squares solution) . Its cumulative frequency distribution is derived from the bootstrap replicate estimates $\hat{\eta}^b$. The replicates are computed by applying the estimation formulae to bootstrap samples. Nonparametric bootstrap replications are obtained when bootstrap samples are generated by random sampling with replacement from the observed data. Here we generated model-conditioned bootstrap replications, which are obtained by sampling with replacement from all the observed abundance index residuals (nonparametric) and adding these to the model predicted values for the abundance indices. As the residuals were weighted by the inverse standard error at each age in the minimization objective function, the resampled weighted residuals used to construct the bootstrap samples were multiplied by the standard error before being added to the predicted index values. For the percentile method, the confidence distribution of the interest parameter is defined as the proportion of bootstrap replicates, $\hat{\eta}^b$, less than or equal to that value, $\hat{\Omega}(x) = \text{Prob}\{\hat{\eta} \leq x\} = \dfrac{\#\{\hat{\eta}^b \leq x\}}{B}$.

The bias-corrected percentile method of Efron (1982), improves on the percentile method by adjusting for differences between the median of the bootstrap percentile density function and the estimate obtained with the original data sample. The confidence distribution of the interest parameter is obtained with the bias-corrected percentile method by constructing the paired values $\left(\hat{\eta}^b_{BC}, \alpha \right)$. The α are the respective probability levels equal to $1/B, 2/B, 3/B, \cdots, B-1/B$ where B is the total number of bootstrap replicates. For each α, calculate the bias adjusted quantity, $\hat{\eta}^b_{BC} = \hat{\Omega}^{-1}(\Phi(2z_0 + z_\alpha))$. Here, Φ is the cumulative distribution function of a standard normal variate, $z_\alpha = \Phi^{-1}(\alpha)$ and $z_0 = \Phi^{-1}\left(\hat{\Omega}(\hat{\eta}) \right)$. The term z_0 achieves the bias adjustment. The notation $\hat{\Omega}^{-1}(\ )$ or $\Phi^{-1}(\ )$ is used to represent the inverse distribution function, i.e. the critical value corresponding to the specified probability level. Note that computations are not carried out for $\alpha = B/B$ because $z_\alpha = \Phi^{-1}(\alpha = 1)$ is not defined.

Bayes - ADAPT

The priors for the ADAPT analysis are assumed to be:

Parameter	Prior distribution
$\{\theta_{26,a} : a = 2,3,..,15\}$	U[-∞, ∞]

$\{\kappa_a : a = 1,2..,15\}$	$U[-\infty, \infty]$
σ_a^2	$\propto \frac{1}{\sigma_a^2}$

XSA Specification

The design of XSA was intended to implement a q-constrained VPA estimation method. The default algorithm assumes that index catchability for the oldest age is equal to that for the next oldest age. For this study, the algorithm was modified to estimate catchability for all ages independently. XSA also estimates the survivors for all year-classes by default. However, an F-constrained VPA can be mimicked through the shrinkage option. Shrinkage to the mean F was applied to the oldest age in all years other than the terminal year. The weighting for F shrinkage was derived assuming a standard error of 0.01.

Observations from the CPUE series were weighted using inverse variances obtained for each age. The variances were derived from the standard errors of the log catchability which are assumed to be log normal. In addition the standard errors were assumed to have a minimum value of 0.3.

Delta-XSA

Numbers at age in the last data year are assumed to have a log normal distribution i.e.

$$Ln\ Nmc \sim Ln\ N + D\ (0,\sigma^2)$$

where N is the XSA estimate of the survivors for a cohort. $D(0,\sigma^2)$ represents the normal distribution with unit mean and variance σ^2 as derived from the XSA model. Within XSA two estimates of standard error are identified - "internal" and "external". To be conservative the larger of the two values at each age was adopted as the estimate of variance. Covariances are ignored.

To estimate statistics of interest draws are made from the distributions of N. These Ns are then used to solve the catch equation using the given catches and natural mortalities. This ensures that fishing mortality is consistent with the observed catches. The parameters of interest, $F_{0.1}$, SSB in the last year and the TAC in the year after the last data year are then derived using these simulated Ns and Fs.

NPB-XSA

The nonparametric bootstrap of XSA models uncertainty in the CPUE indices and hence other quantities of interest by performing a nonparametric bootstrap of the relationship between CPUE and population size. CPUE is transformed to relate the population abundance during the time at which the catch was taken to the population abundance at the beginning of the year. Time series weighting was not used so that the residuals were selected with equal probability.

The bootstrap algorithm

1) Run XSA and estimate log catchability
2) Sample by year from the "log catchability residuals" within age and fleet with replacement
3) Generate new CPUE indices
4) Re-run XSA and generate statistics of interest

5) Repeat steps 2)-5)

RESULTS AND CONCLUSIONS

Figures 3-12 show the results obtained for the different procedures. Referring to Figure 2, the following patterns were observed in the results:

Patterns in the results.A "+" denotes good performance; a "B" indicates that the parameter is overestimated - refer to Figure 2 for diagnostic patterns

Method	SSB	F0.1	TAC
Delta-ADAPT	B	+	B
Delta-ADAPT-BC	+	+	+
NPB-ADAPT	B	+	B
NPB-ADAPT-BC	+	+	+
PB-ADAPT	B	+	B
Bayes-ADAPT	B	+	B
Delta-XSA	B	+	B
NPB-XSA	B	+	B

The following table provides an indication of the accuracy with which the performance statistics were estimated. Each entry is the number of simulations out of 100 for which the true value was smaller than the estimated 50th percentile (i.e. entries equal to 50 indicate accurate results; higher values indicate over-estimation).

Method	SSB	F0.1	TAC
Delta-ADAPT	56	49	61
Delta-ADAPT-BC	50	52	53
NPB-ADAPT	71	45	71
NPB-ADAPT-BC	51	51	54
PB-ADAPT	68	52	74
Bayes-ADAPT	57	54	66
Delta-XSA	75	56	77
NPB-XSA	59	59	63

For these VPA-based methods, inference statements about F0.1 tended to be more accurate than those for SSB or TAC, which tended to be overestimated (except for bias corrected NPB-ADAPT and bias corrected Delta-ADAPT applications).

The Bias Corrected Delta-ADAPT and Bias Corrected NPB-ADAPT applications performed better than the others. Bias correction, when it can be applied, tends to help improve the accuracy of inference statements.

It is evident that accurate inference statements about some parameters can be obtained with the approaches examined here.

Methods with bias correction appeared to perform well, but a troubling result is that the other methods overestimated SSB and TAC. These methods are in quite widespread use for advisory purposes. The results tabulated above imply that, for these data, choosing a TAC corresponding to the median, for example, could result in fishing mortality levels that exceed the target in about 7 out of 10 cases (instead of the 50% intended by choosing the median). We do not know the degree to which this result can be generalized because we only examined one simulated population with a given trajectory and a given set of observation errors. It may be informative to repeat this exercise with different population trajectories to see what patterns emerge. Nevertheless, these simulations suggest that the magnitude of the bias can be of concern.

As mentioned earlier, we conducted these tests ensuring that the structural models in the assessment procedures were specified in the same way as the simulated population was generated. We believe that the issue of model mis-specification needs to be confronted in studies of the type we have conducted here. However, the model components that can be mis-specified are numerous, making this an onerous exercise.

ACKNOWLEDGMENTS

REFERENCES

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

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.

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.

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.

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

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

Lewy, P. and Nielsen, A. (2000). A Bayesian approach to fish stock assessment of North Sea plaice using Markov Chain Monte Carlo and graphical models. Working Paper for the EU concerted action FAIR PL98-4231 project meeting in Madrid, February 2000.

Mesnil, B. (1995) Analyse de risque en évaluation de stocks et gestion de pêcheries. Utilisation des méthodes de Monte Carlo. In: Les recherches françaises en évaluation quantitative et modélisation des ressources et des systèmes halieutiques (Actes du 1er Forum Halieumétrique, ENSA, Rennes 29 juin - 1er juillet 1993) D. Gascuel, D., J.L. Durand and A. Fonteneau, eds. ORSTOM, Colloques et séminaires, Rennes. 105-125. [In French.]

Punt, A.E. and Hilborn, R. (1997) Fisheries stock assessment and decision analysis: A review of the Bayesian approach. Reviews in Fish Biology and Fisheries 7, 35-63.

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

Restrepo, V.R., Hoenig, J.M., Powers, J.E., Baird, J.W. and Turner, S.C. (1992) A simple simulation approach to risk and cost analysis, with applications to swordfish and cod fisheries. Fishery Bulletin, US. 90, 736-748.

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.
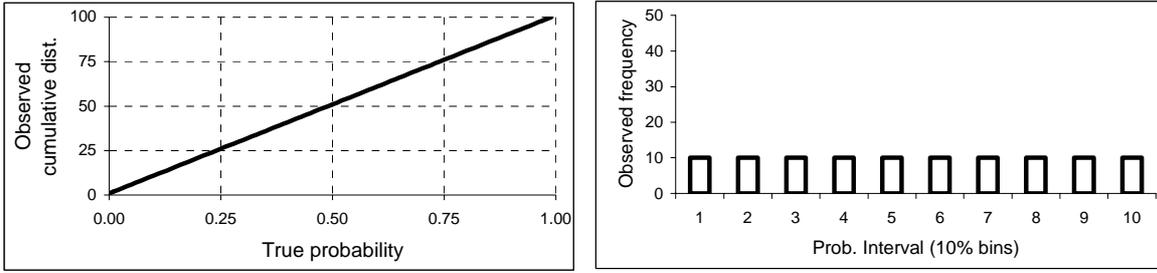
Figure 1. Example characterization of the results for a model with perfect accuracy. The left panel shows the cumulative number of simulation realizations resulting in a given true probability statement; the right panel shows the same information categorized in 10 intervals of equal size.
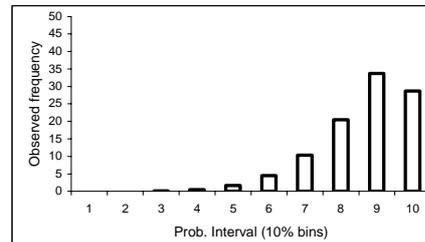

True μ **higher** than μ estimate
True σ **equal** to σ estimate

A

True μ **smaller** than μ estimate
True σ **equal** to σ estimate

B

True μ **equal** to μ estimate
True σ **higher** than σ estimate

C

True μ **equal** to μ estimate
True σ **smaller** than σ estimate

D
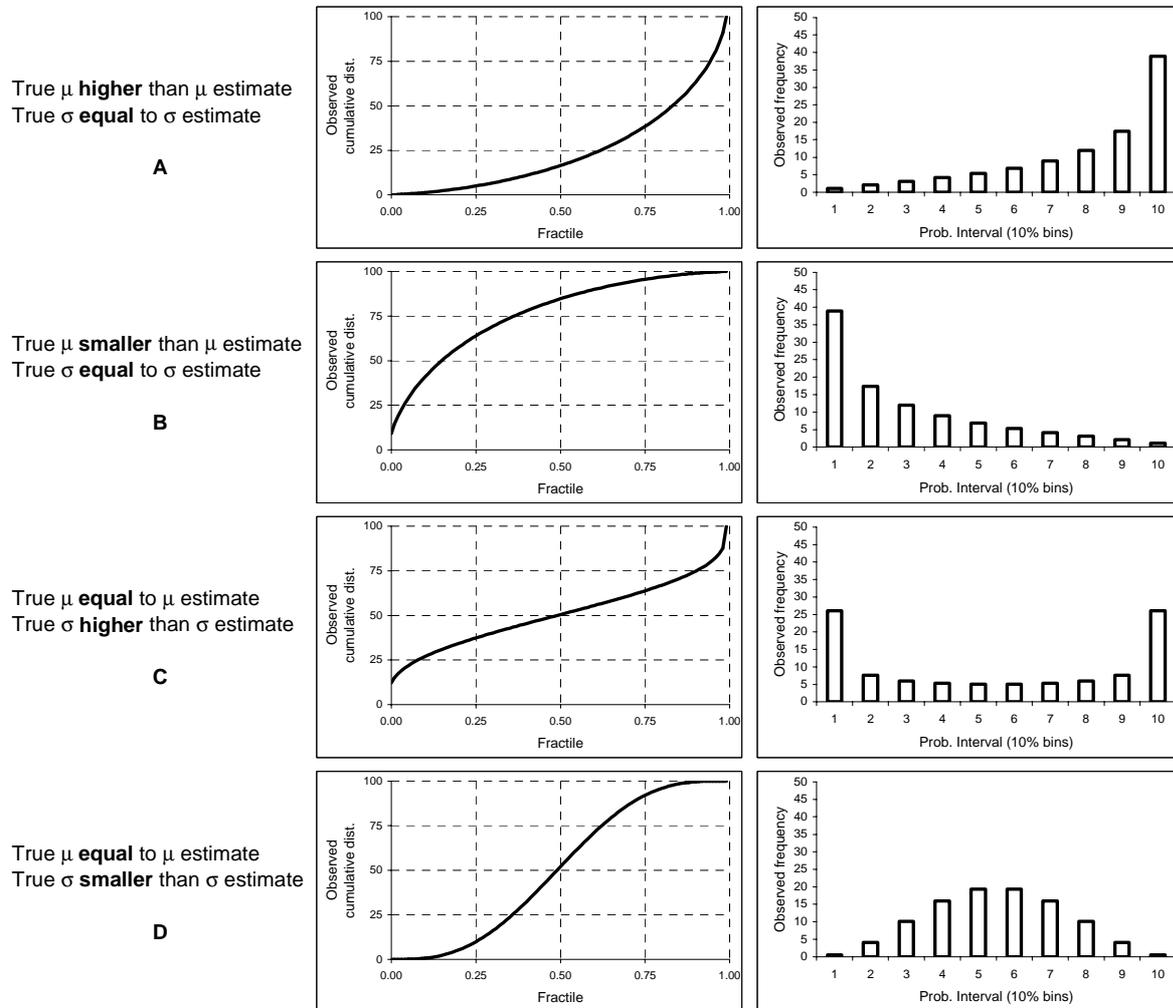
Figure 2 (a). Example simulation results for methods that do not perform well. The text on the left of each pair of graphs explains the cause of inaccuracy, in terms of the location (μ) and dispersion (σ) parameters.
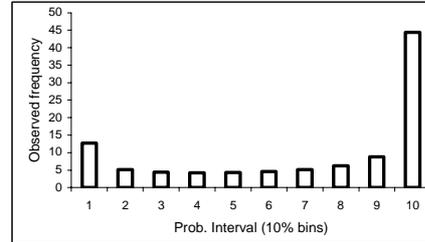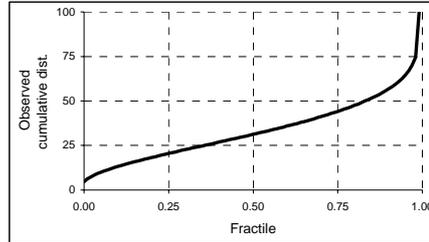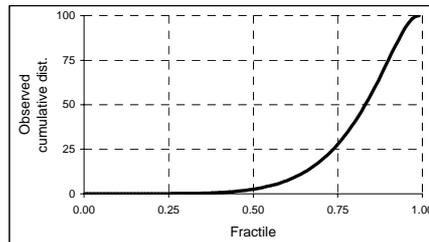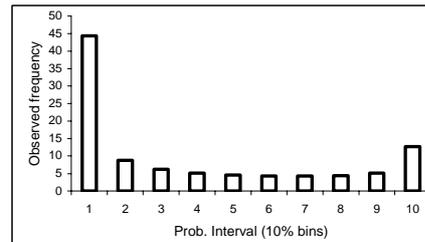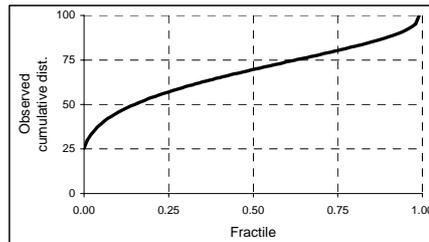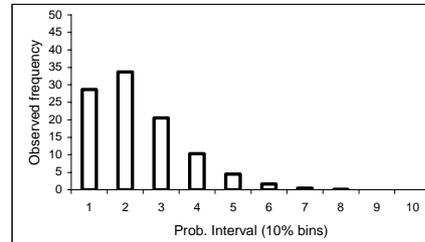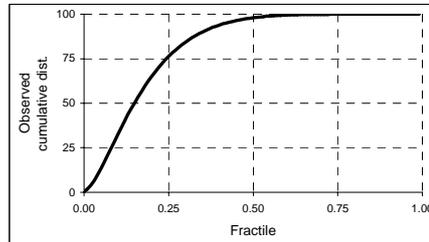
Figure 2 (b). Example simulation results for methods that do not perform well. The text on the left of each pair of graphs explains the cause of inaccuracy , in terms of the location (μ) and dispersion (σ) parameters.
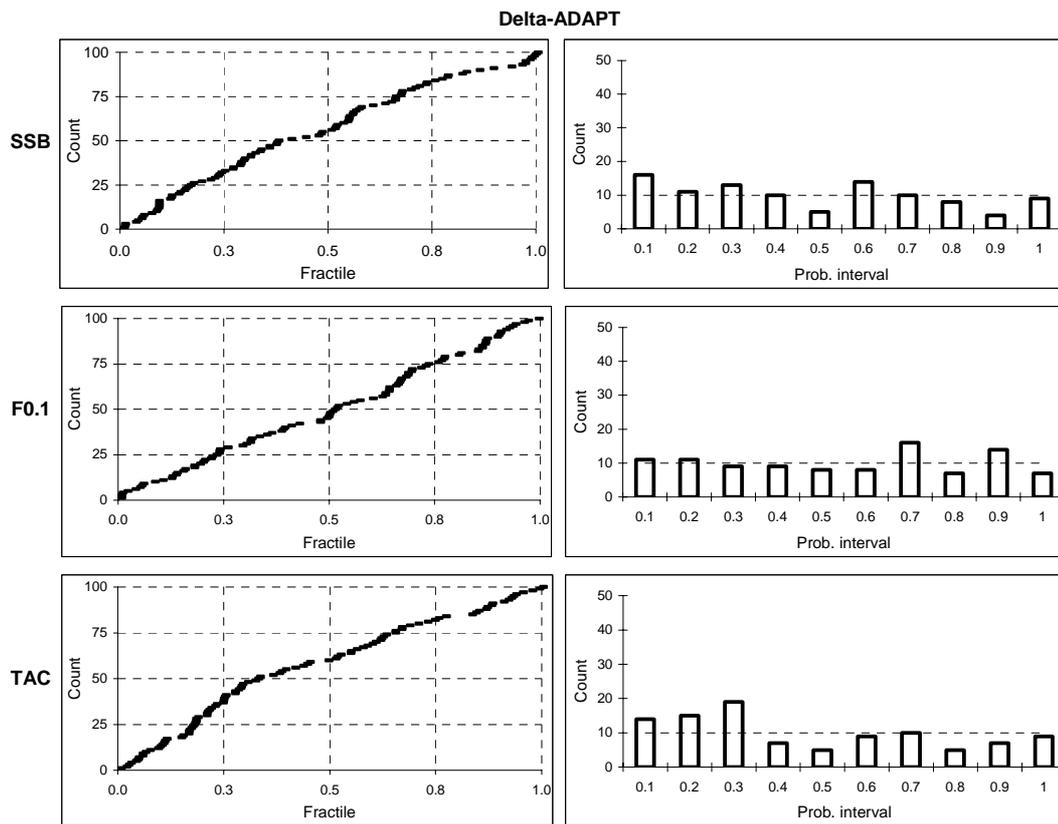
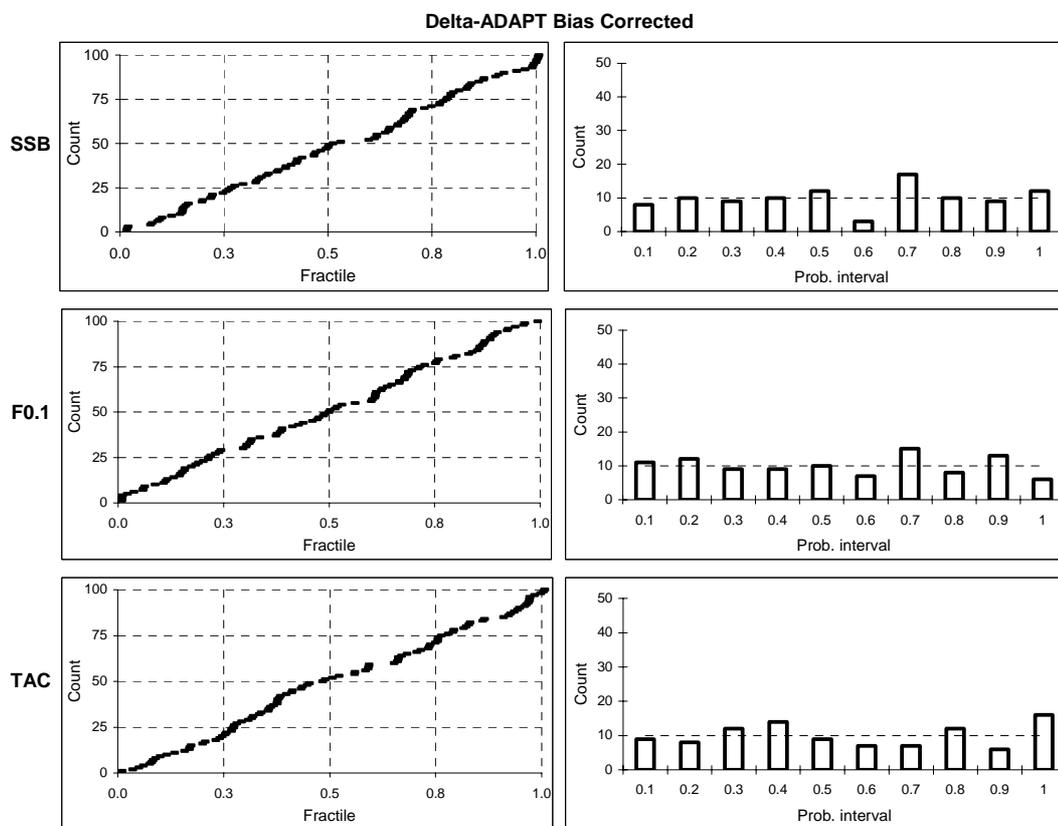**Figure 3.** Simulation results for Delta-ADAPT

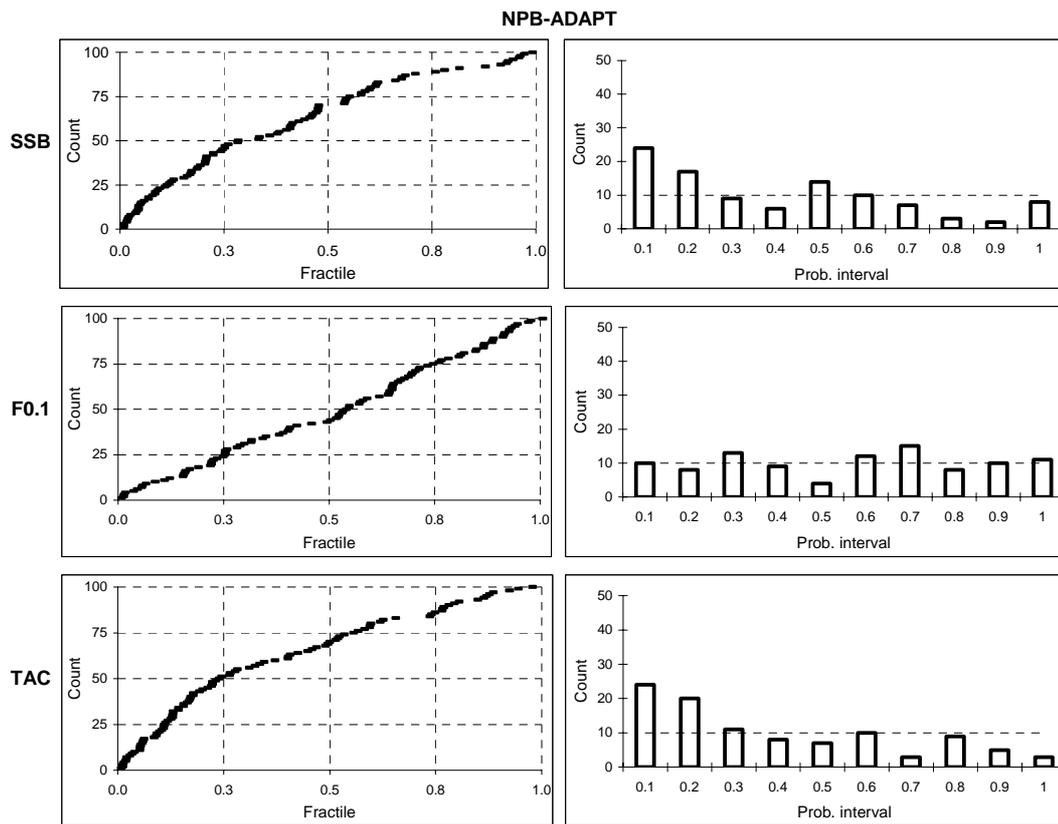**Figure 4**. Simulation results for Delta-ADAPT, bias-corrected.
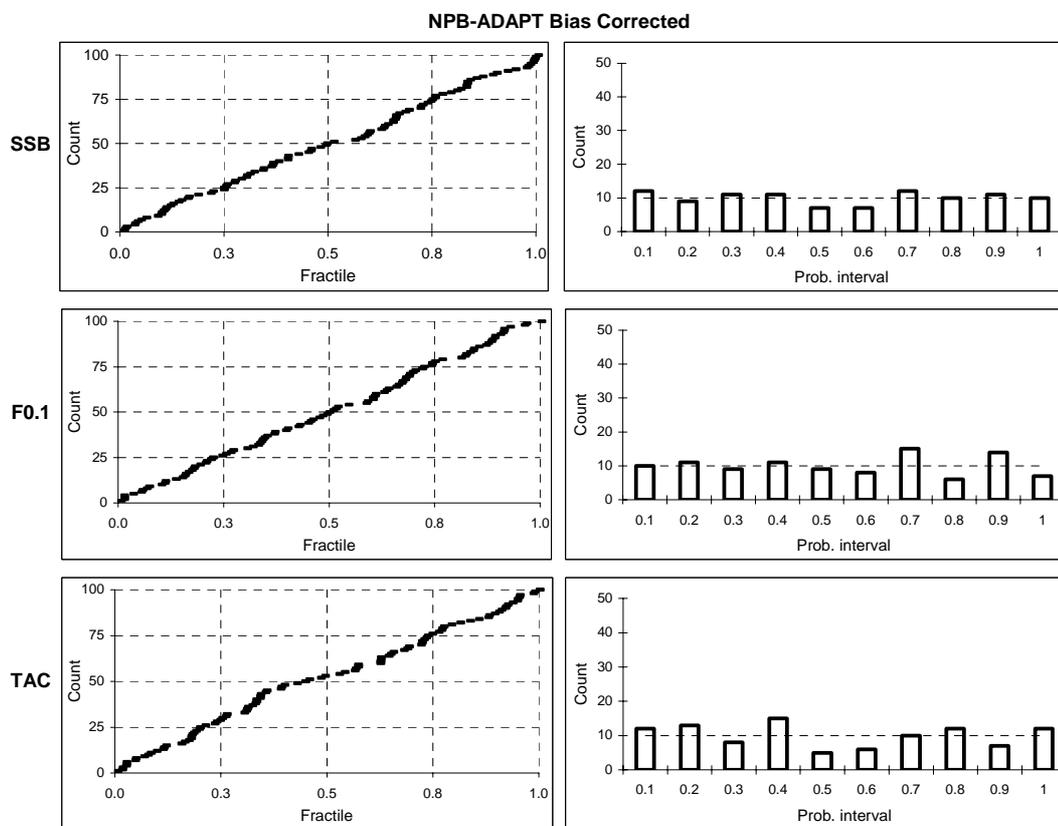
Figure 5. Simulation results for NPB-ADAPT.
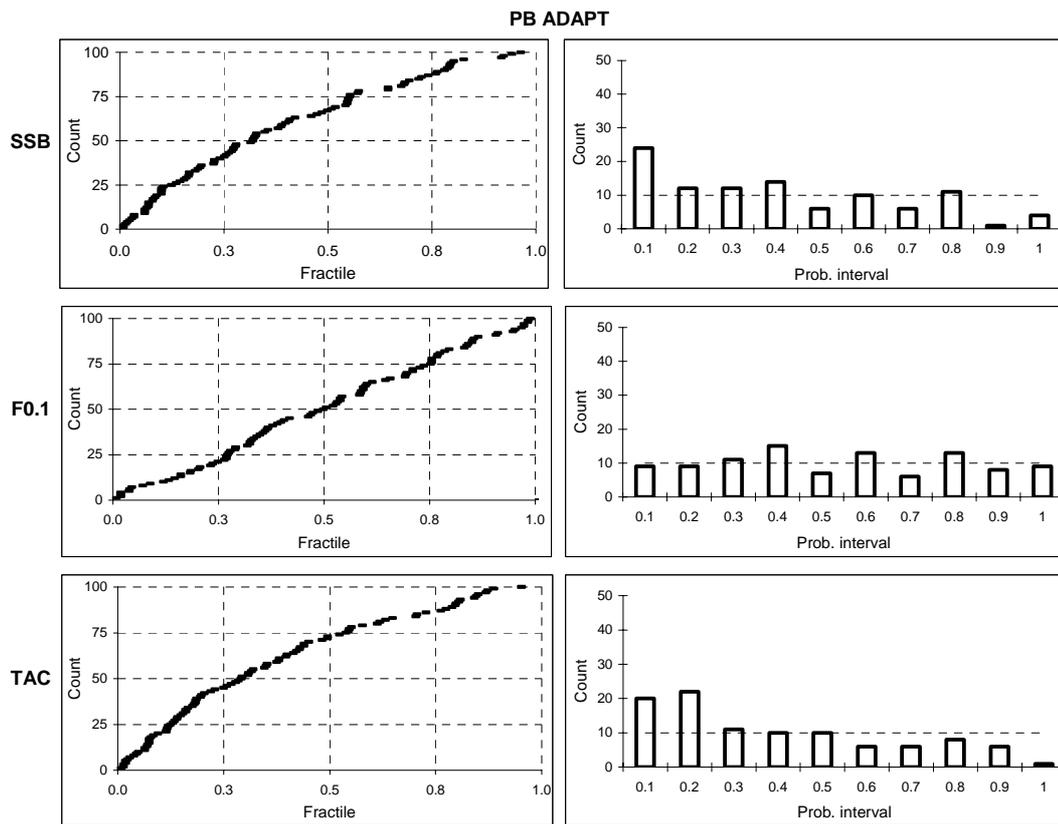

Figure 6. Simulation results for NPB-ADAPT, bias-corrected.

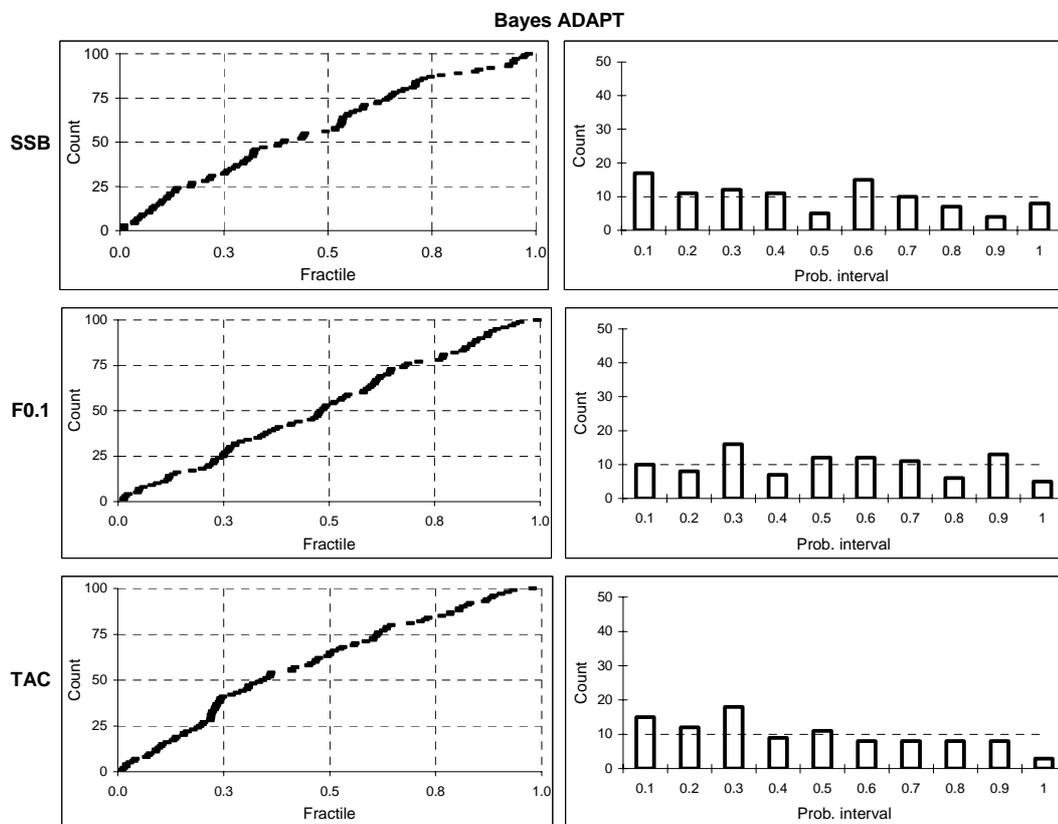**Figure 7**. Simulation results for PB-ADAPT
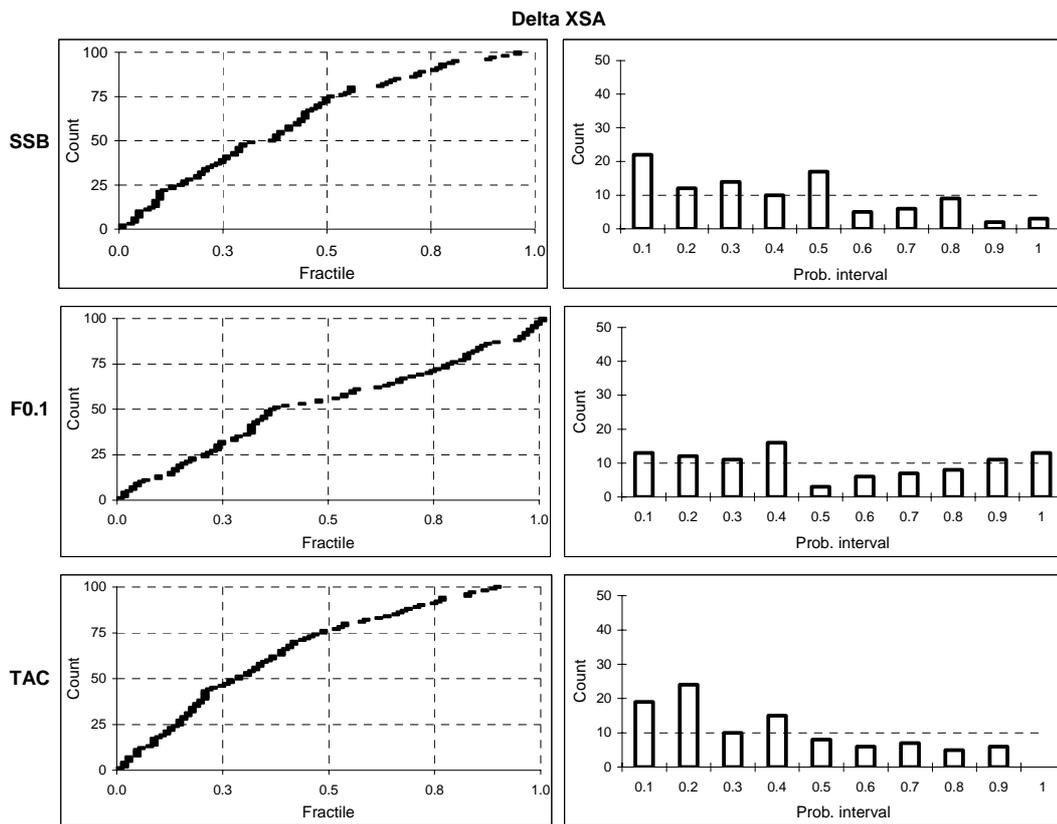

Figure 8. Simulation results for Bayes-ADAPT.
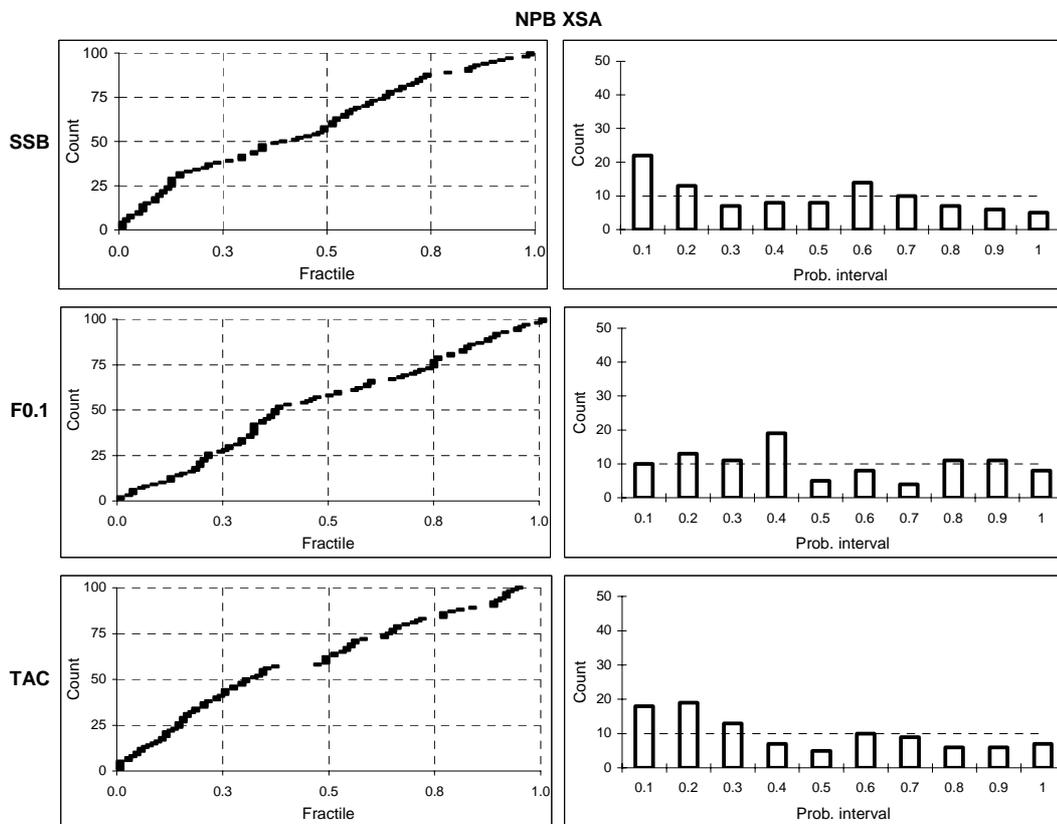
Figure 9. Simulation results for Delta-XSA.

Figure 10. Simulation results for NPB-XSA.