Not to be Cited Without Prior Reference to the Authors

ICES 1993 C.M. D:12 Statistics
Comm. Session P

The Eloquent Shrug:
Expressing Uncertainty and Risk in Stock Assessments

A. A. Rosenberg
National Marine Fisheries Service
Woods Hole, MA, USA

and

V. Restrepo
University of Miami
Miami, FL, USA

This paper explores methods for expressing uncertainty in the scientific advice on current and projected resource status. We have conducted simulation experiments incorporating variability in the stock-recruitment relationship and measurement error in catch and abundance estimates for stocks with different life histories, exploitation and information bases. Stock assessments are simulated for every projection year and TAC's are chosen based on management strategies given by the estimated target biological reference point and the probability that the TAC will achieve its target.

Using these simulation results, where there is a "perceived" population as well as a "true" underlying population, we consider how the choice of management strategy interacts with the uncertainty in the assessment of current status. We explore possible measures of risk to the resource and the fishery for different strategies. In discussing risk we consider related issues of defining biomass thresholds, statistical power and Type I error. We then discuss how these sorts of calculations may be incorporated into assessment advice on a routine basis.

INTRODUCTION

As stock assessment methodology has moved toward statistical procedures, increasing attention has been focussed on estimating the uncertainty in assessment advice (e.g. Shepherd 1991, Smith et al 1993, Anon 1993). It has become routine in many assessments to estimate the variance of the estimated stock abundance or fishing mortality rate, but in many cases this additional information is primarily used by scientists as a diagnostic for the analysis. Expressing assessment uncertainty to managers is another matter however, and there is, as yet, no standard approach in the presentation of advice with respect to uncertainty and risk. Here, we investigate the role of uncertainty in the provision of management advice through simulation studies. We present a number of different measures of the projected status of the resource and the implications of managers adopting different approaches to the inherent uncertainty in the assessments. Then, we discuss how such information may be incorporated into scientific advice for managers.

In broad outline, we are seeking to address two fundamental questions with our simulations; 1) how can a given management strategy be evaluated on a year to year basis given uncertainty in knowledge of resource status ? and 2) what is the benefit of reducing uncertainty as information accumulates through time or additional research ? We consider these questions by generating model fish populations roughly similar to a Northwest Atlantic cod stock and apply an $F_{0.1}$ (Gulland and Boerema 1973) or an F_{med} (Sissenwine and Shepherd 1987) harvest strategy over a 15 year period. Different levels of uncertainty are introduced into the process for comparison. We carry through the full assessment process using the ADAPT procedure (Gavaris 1988; Conser and Powers 1989) in each year of the simulation, projecting one year ahead to set catch limits, then updating the estimates of stock status. We then compare performance using a suite of measures evaluated on the "perceived", i.e., estimated, population and the true, simulated population. This distinction is crucial in our view. We wish to know how the management measures have performed for the underlying population. But, equally or perhaps more important is the ability to detect what is happening given uncertainty and a particular management strategy.

Uncertainty and Risk

The analysis of uncertainty and risk assessment are often taken to mean the same thing in fisheries. In decision theory, this is not the case. The term risk means the expected loss of benefits from the resource, where the benefits are defined with respect to some utility or loss function which expresses the value of possible outcomes such as yield or recruitment. In practice, the term risk is often used to describe some subjective notion of danger to the resource base, such as the risk of spawning biomass falling below a specified level, or the risk of recruitment failure (Restrepo et al. 1992; Francis 1992). It would certainly be desirable to use the more formal approach of decision theory, but this would require managers making clear the appropriate utility or loss functions to be applied.

Uncertainty arises from an imprecise knowledge of the state of nature. At least five types of uncertainty can be distinguished:

1) measurement error is the error in the observed quantities such as the catch or biological parameters.

2) process error is the underlying stochasticity in the population dynamics such as the variability in recruitment.

3) model error is the misspecification of model structure.

4) estimation error can result from any of the above uncertainties and is the inaccuracy and imprecision in the estimated population parameters such as stock abundance or fishing mortality rate.

5) implementation error results from variability in the resulting implementation of a management policy, i.e., inability to exactly achieve a target harvest strategy.

In the simulation studies presented here, we incorporate measurement error, process error and estimation error. We evaluate risk in comparison to simple utility functions related to yield and recruitment. In addition, we consider the more usual notion of risk as the chance of low stock biomass or poor recruitment and calculate the statistical power and a form of Type I error in the assessment advice as an important component of uncertainty for managers.

SIMULATION STRUCTURE

We generated model fish populations using standard fishery population dynamics equations with stochastic, lognormally distributed recruitment with mean recruitment at a given spawning biomass level governed by a Beverton and Holt stock and recruitment relationship and a constant coefficient of variation as an input parameter. Other input parameters were the rate of natural mortality, the maturity ogive, weight at age, and the partial recruitment at age to the fishery. For each population we also input an exploitation history, either overexploited or underexploited. In the overexploited case, the simulated population was harvested at an $F_{0.1}$ rate for 50 years. Then the mortality rate was increased over a 10 year period to a fishing mortality rate corresponding to the slope at the origin of the underlying stock recruitment curve (note because of stochasticity this harvest rate did not result in rapid extinction of the population), which was maintained for a further 10 years before the first projection year. In the underexploited case, harvesting at $F_{0.1}$ was maintained up until the first projection year.

As noted above, an assessment was conducted each simulation year starting with the first projection year. These assessments used the ADAPT method of least squares fitting of the catch at age matrix and research survey indices at age. Measurement error was added to the catch at age matrix from a lognormal distribution with input coefficient of variation. Similarly, lognormal errors were added to the survey indices. A separable VPA was performed on the "observed" catch at age matrix to estimate the partial recruitment in the final year for cohorts whose terminal F was not directly estimated as a parameter in the search. Estimates of stock abundance and fishing mortality rates at age from ADAPT gave the "perceived" stock status and perceived reference points, which were re-estimated each year. The partial recruitment vector used to estimate $F_{0.1}$ or F_{med} was the geometric mean of the perceived fishing mortality rates at age for the last three years of data. The perceived reference fishing mortality rate was then applied to the perceived stock to estimate a TAC for the coming year. The true stock was updated using this TAC, a new recruitment value was added and the process repeated.

Simulation experiments were performed for 16 cases in total and in each experiment 100 populations were projected for 15 years each. The random recruitment sequences added to the populations were the same in each experiment to enable comparisons. For each harvest rate strategy, eight experiments were performed, four with overexploited populations and four with underexploited populations. Two of the overexploited and two of the underexploited cases were based on an initial observed data series of 10 years in the first projection year with error C.V.'s of 50% added to the catch and the survey data. The other two were with 20 years of initial data and error C.V.'s of 25%. Of these pairs, in one case the resulting estimate of stock abundance was interpreted cautiously by using the 25th percentile of the distribution of the estimate to calculate the TAC for the coming year. In the other case of each pair, the abundance estimate was viewed optimistically, by using the 75th percentile of the distribution of the estimate.

This design enabled us to consider the benefits of being cautious in interpreting the estimates and the benefits of longer time series or lower error variances in the data for either type of stock or target harvest rate.

SIMULATION RESULTS

To provide a frame of reference for the results, we calculated $F_{0.1}$ as 0.162 and F_{max} as 0.274. for our model populations. B_{msy} for the deterministic yield curve was 50,000 MT and the maximum recruitment on the deterministic stock and recruitment curve was 100,000 fish.

The median fishing mortality rates actually applied to the stocks (the true F's) were similar, between 0.1 and 0.2, for all cases except the overexploited stocks harvested using an F_{med} strategy (Figure 1a). This is simply because F_{med} is calculated from the stock biomass and recruitment levels recently experienced by the population and is a maintenance harvest rate, so will maintain those stock levels. In the overexploited cases, F_{med} gives a fishing mortality rate similar to the recent F the stock experienced, i.e., near the slope of the stock recruitment curve. Note in the Figure 1a that the median harvest rates for the optimistic view of stock status (75% of the distribution of biomass estimates) are always higher than the cautious view.

The perceived harvest rates overestimate the true F when a cautious view of the stock is taken by a substantial amount, nearly double in some cases. In contrast, an optimistic view of the stock abundance results in slight underestimates of F (Figure 1b). This is because using a lower biomass estimate to compute TAC's gives a lower true F compared to the perceived reference point. Longer data series lessen this effect (Figure 1b).

When the F_{med} harvest rate is estimated without error and applied to the stock, the perfect implementation case, it does in fact result in maintenance of the stock at the level at the start of projections. $F_{0.1}$ results in substantial stock rebuilding for the overexploited stocks and maintenance for the underexploited stocks (because it is close to F_{med}. However, using the estimated F_{med}, the overexploited stocks rebuild somewhat by the end of the simulation anyway, more so for the cautious strategies and when the uncertainty is higher in fact. This is because there is a tendency to underestimate stock biomass in the simulated assessments. Using the cautious view of stock abundance, this tendency to underestimate is accentuated and enhances rebuilding. Even so, the average annual true yields are much higher for the underexploited stocks than those overexploited (Figure 2a). The average yield is actually not very sensitive to the degree of uncertainty. The perceived average yield is always slightly less that the actual yield (Figure 2b), but this just results from the lognormal distribution of error in total catch that we used to generate the observations, i.e., the median of a lognormal distribution is less than the mean. Note the short data series highlight this effect.

Another important attribute of fishery performance is the annual variability in yield (Figure 2c). Expected variability in yield, measured by its coefficient of variation, is always less when the uncertainty is lower and is much higher for the overexploited stocks than the underexploited.

We compared the average cumulative yield over the 15 years of projections to both the case where the strategy was implemented perfectly, i.e., there was no measurement error so abundance was known exactly, and to a population harvested at the status quo fishing mortality rate at the start of the projections. The former gives a measure of the foregone yield due to measurement error. The latter is a measure of the foregone yield from taking different harvest strategies, here the choice of $F_{0.1}$ or F_{med} combined with a cautious or optimistic view of management. Note that foregone yield in either case is defined as the respective perfect implementation or status quo yield minus the true yield obtained in the simulated case. Therefore, positive foregone yields mean that the perfect implementation or status quo yield was higher than for a particular combination of measurement error and harvest strategy.

For the underexploited stocks, the foregone yield calculations indicate that it does not pay to be extra cautious (in terms of the estimate of stock abundance) and, for lightly exploited stocks, one can afford to be optimistic (Figure 3). If the stock is heavily exploited an $F_{0.1}$ is an advantageous strategy compared to the status quo of course because it gives a lower fishing mortality rate. The true yield under F_{med} is also higher than the status quo because of the tendency to underestimate F noted above. Because higher uncertainty implies wider confidence limits on the estimated biomass, a shorter time series and cautious strategy give large negative foregone yields, i.e., fishing mortality rates are lowered even more. The longer time series with lower uncertainty always provide yields closer to the perfect implementation case of course.

Yield is one measure of stock productivity, but recruitment is also an important measure in its own right. One goal of management may be to maintain long term viability, i.e., prevent recruitment overfishing, regardless of whether the yield is the largest obtainable. We examined expected recruitment for each case (Figure 4a). Underexploited stocks had much higher expected recruitment than overexploited stocks because of the shape of the stock recruitment relationship built into the simulations. In every case, recruitment was slightly higher using a cautious view of abundance and differences between cautious and optimistic were larger for cases where the uncertainty was higher (short time series).

The perceived recruitment was higher than the true simulated recruitment for the overexploited stocks, sometimes by as much as 60% (Figure 4b). This has implications beyond the strict comparisons made here. Overestimating recruitment if the projections were made more than one year ahead to guide management, would suggest the stock is producing at a higher level than is actually the case. Note that the runs with greater uncertainty overestimate recruitment more.

The Risk of Recruitment Overfishing

Maintaining stock productivity implies maintaining the production of new recruits and therefore maintaining a sufficient spawning stock to keep the probability of good recruitment high. Determining what constitutes "good" recruitment is an arbitrary, though hopefully sensible decision and, unless there is clear evidence for depensatory recruitment, the choice of a spawning stock abundance to maintain a high chance of good recruitment is somewhat arbitrary. Here, we choose to define good recruitment as greater than or equal to half of the maximum recruitment from the underlying relationship between stock and recruitment. The threshold for spawning biomass therefore is the biomass that is expected to give recruitment at one half the maximum of the deterministic relationship (Mace in press, Myers et al. in press). For the Beverton-Holt relationship written as $R = aS/(1+S/K)$ this is given by the parameter K.

We calculated the probability that the true simulated recruitment was less than half the maximum of the underlying curve and the probability that spawning biomass was less than the level that would is expected to produce half the maximum recruitment. The results were similar and only the spawning biomass threshold results are described below.

For the underexploited cases, spawning biomass never fell below K. Fishing at F_{med} on an overexploited stock kept the stock below the threshold and recruitment consequently low (Figure 5a). A cautious view of abundance only slightly ameliorates this is situation, but the $F_{0.1}$ reduces this measure of risk to the stock to between 10 and 20%.

Of course, in practice, one would not know these probabilities or the true, underlying stock and recruitment curve. This could only be estimated from the data and the estimated spawning biomass compared to the estimated threshold level would be used to judge if the stock was at risk of reduced recruitment. We compared this perceived risk to the stock to the true simulated population and calculated statistical power and the chance of Type I error. Here, power is the probability of perceiving that the stock is below the threshold when in fact it is. Type I error is the probability of perceiving that the stock is below the threshold when in fact it is not. In these tests we used an alpha level of 0.2, i.e., if the estimated threshold fell above the 80th percentile of the distribution of the estimated spawning biomass the stock was considered below the threshold. This is also another sort of Type I error, from drawing an incorrect conclusion due to the alpha level chosen, fixed at 20%. The additional error comes from a combination of misestimation of the abundance and the threshold.

To calculate the threshold we used two methods, that due to Serebryakov (1992; Shepherd 1992) and simply fitting a Beverton Holt curve to the observed data for each year and using estimated K as the threshold. Note that the perceived threshold moves each year as the assessment is updated. The Serebryakov method performed poorly and was very unstable as biomass changed (Myers et al. in press b) and is not discussed further. The power for the fitted curves was much higher for the longer data series with little difference between the cautious and optimistic views of abundance (Figure 5b). The Type I error, however, was always higher for the optimistic harvest strategies and lower for the longer data series (Figure 5c). We examined the time series of threshold estimates compared to the true K of the underlying relationship. There is a clear pattern of learning as data accumulates through time as expected.

DISCUSSION

Our simulations provide some general results on the importance of uncertainty in the assessments. The benefits of reduced uncertainty in the form of longer time series and lower measurement error show up most clearly in the expected variability of yield and the power and Type I error for detecting risk to stock with respect to recruitment. These are important quantities for any management decision and should where possible be calculated as a component of the scientific advice as discussed below.

The choice between harvest strategies is less clear cut. There seems little advantage to choosing a cautious view of abundance if an already cautious strategy such as $F_{0.1}$ is chosen. The benefits of caution here are with respect to risk of the stock going below a threshold level, statistical power and Type I error. However, these are powerful arguments, since there is generally little to lose from added caution. The F_{med} strategy appears to be outperformed by $F_{0.1}$ as a harvest rule and is probably more appropriate as a harvest threshold than as a target. It is likely intermediate strategies (Pelletier and Laurec 1991) or a more complicated harvest control law (Rosenberg 1993) would outperform either of these simple strategies.

Perhaps the most important point of the work presented here is the various measures of the performance of different harvest strategies. In our view it would be useful to calculate routinely for many assessments measures such as expected yield and recruitment, expected variability in yield, probability of going below a biomass threshold or risk to the stock and the power and Type I error of the estimates for detecting such effects. In the context of a real simulation this would require some simulations studies using the available data to bootstrap projections into the future. While we have simulated a full updating procedure here, this is not necessarily needed for a real assessment. One approach would be to resample or Monte Carlo recruitment estimates from observed recruitments or their estimated distributions for different stock sizes as a basis for the projections under different harvest strategies. Estimates of measurement error in the assessments should also be bootstrapped or used in Monte Carlo experiments. The measures are then calculated over medium term projection as we have done here.

The power analyses done here are somewhat more difficult, but can still be performed as an important component of the advice. A fitted stock and recruitment curve for the observed data is needed as the basis for the analysis. This is taken as the "true" curve as in our studies. New sets of stock recruitment data could then be generated via bootstrapping or Monte Carlo methods and power and Type I error computed as the number of times the stock appeared to be below the threshold (1/2 maximum recruitment of a fitted curve) in the generated data sets when it was in the original data set. An additional difficulty arises due to the range of observations of stock biomass. A very different impression of the relationship between stock and recruitment often results from stocks which have undergone different exploitation histories. Because of this, it may be useful to hypothesize a number of different curves for trial in the power analysis. However, the question of relevance is how sensitive is the conclusion concerning the status of the resource vis a vis one or more biological reference points, i.e., do we get a different answer using different realizations generated from the distribution of the parameters.

Clearly this is not a complete power analysis, but at least will guide managers as to the ability of the assessment methods to warn them of danger to the stock. Similarly, the Type I error warns of the chances of taking difficult and costly actions when they are not warranted.

While these simulations may appear to add a large burden to the assessment process, we note that many of the needed quantities are already output in many assessments (measurement error variances, projections under different strategies). We simply argue for completing the process and carrying through the estimates of uncertainty to the management advice itself.

Literature Cited

Anon. 1993. Report of the working group on methods of fish stock assessment. ICES C.M. 1993 Assess: 12.

Conser, R. J. and J. E. Powers. 1989. Extensions of the ADAPT VPA tuning method designed to facilitate assessment work on tuna and swordfish stocks. ICCAT Working Doc. SCRS/89/43.

Francis, R. C. C. 1991. Risk analysis in fishery management. NAFO Sci Coun. Studies 16: 143-148.

Gavaris, S. 1988. An adaptive framework for the estimation of population size. CAFSAC Res. Doc. 88/29.

Gulland, J. A. and K. Boerema. 1973. Scientific advice on catch levels. Fish. Bull. U.S. 71:325-335.

Mace, P.M. in press. Relationships between common biological reference points used as thresholds and targets of fisheries management strategies. Can. J. Fish. Aquat. Sci. 50.

Myers, R. A., A. A. Rosenberg, P. M. Mace, N. Barrowman and V. R. Restrepo. in press. In search of thresholds for recruitment overfishing. ICES J. Mar. Sci.

Pelletier, D. and A. Laurec. 1991. Toward more efficient TAC policies with error prone data.NAFO Sci Coun. Studies 16: 153-163.

Restrepo, V. R., J. M. Hoenig, J. E. Powers, J. W. Baird and S. C. Turner. 1992. A simple simulation approach to risk and cost analysis, with applications to swordfish and cod fisheries. Fish Bull. U.S. 90: 736-748.

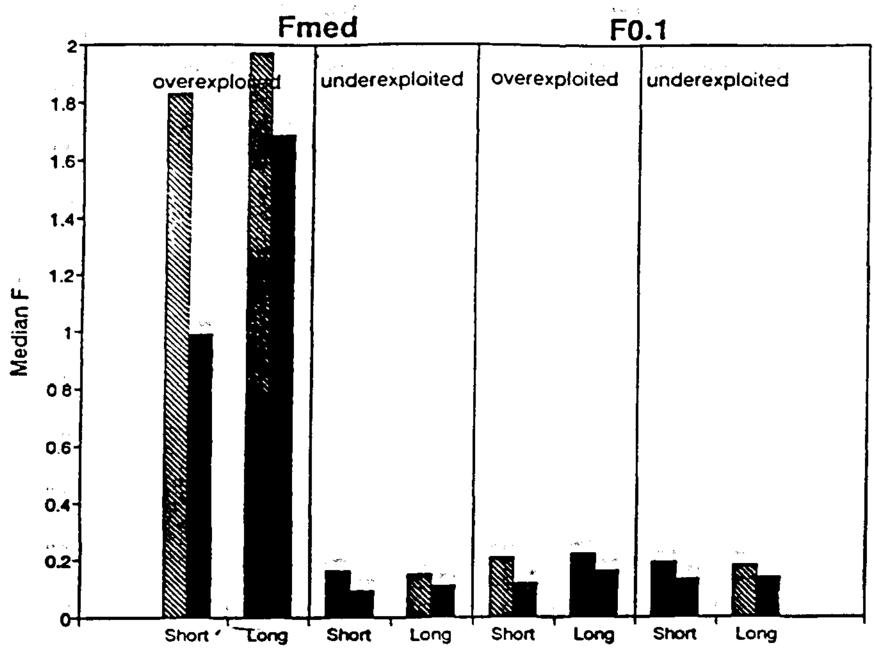
Rosenberg, A. A. (ed.) 1993. Defining overfishing- defining stock rebuilding: report of the second annual national stock assessment workshop. NOAA Tech. Memo. NMFS-F/SPO 8: 68p.

Shepherd, J. G. (ed.). 1991. management under uncertainties. NAFO Sci Coun. Studies 16.

Sissenwine, M. P. and J. G. Shepherd. 1987. An alternative perspective on recruitment overfishing and biological reference points. Can. J. Fish. Aquat. Sci. 44: 913-918.

Smith, S. J., J. J. Hunt and D. Rivard. (eds.) 1993. Risk evaluation and biological reference points for fisheries management. Can. Spec. Pub. Fish. Aquat. Sci. 120.

A.



B.

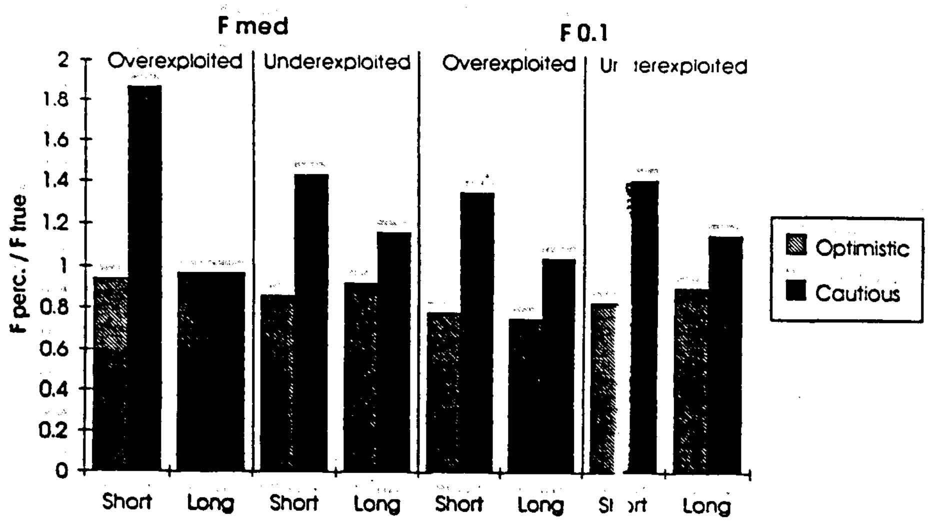
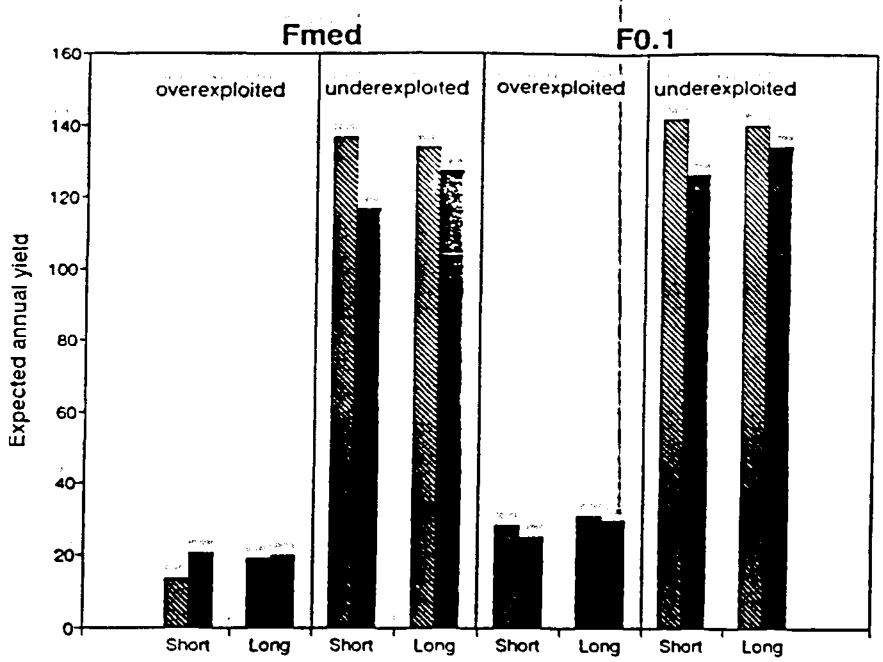


Figure 1: A. Median true F in the simulations case by case. The solid bars are the runs where the 75th percentile of the estimated biomass distribution was used to set the TAC. The hatched lines used the 25th percentile.
B: The ratio of the perceived F to the true F in the simulations case by case.

A.



B.

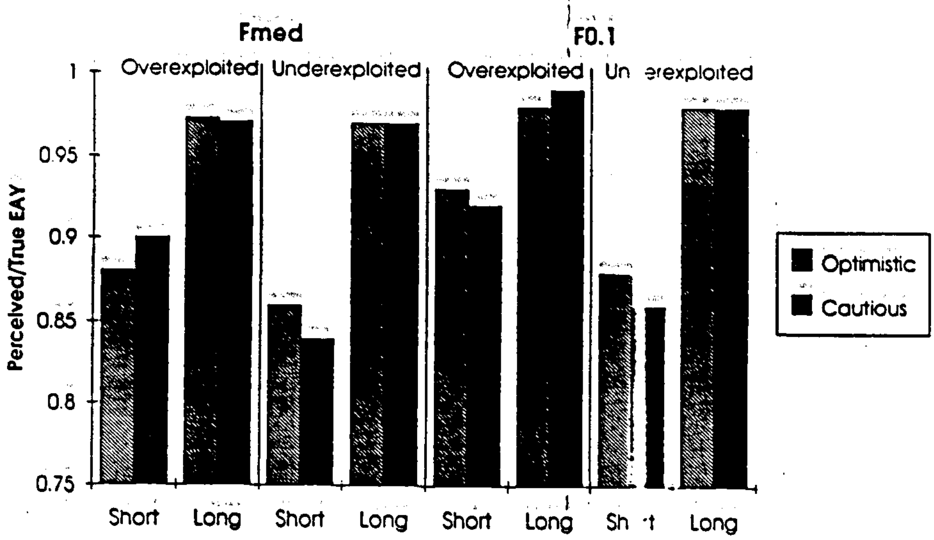
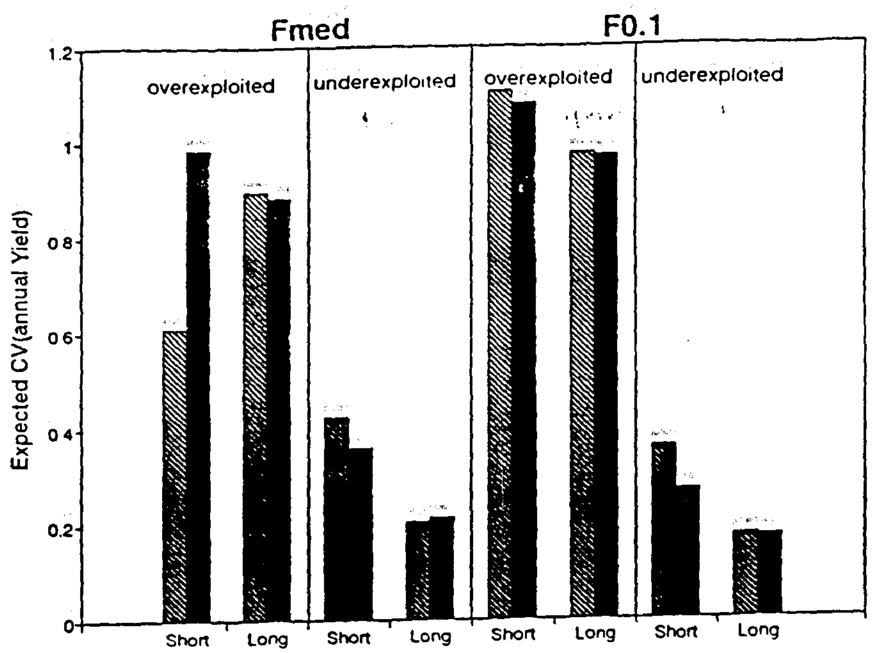


Figure 2: A. The expected annual yield case by case averaged across years and realizations. B. the perceived expected annual yield as a fraction of the true.

A.



B.

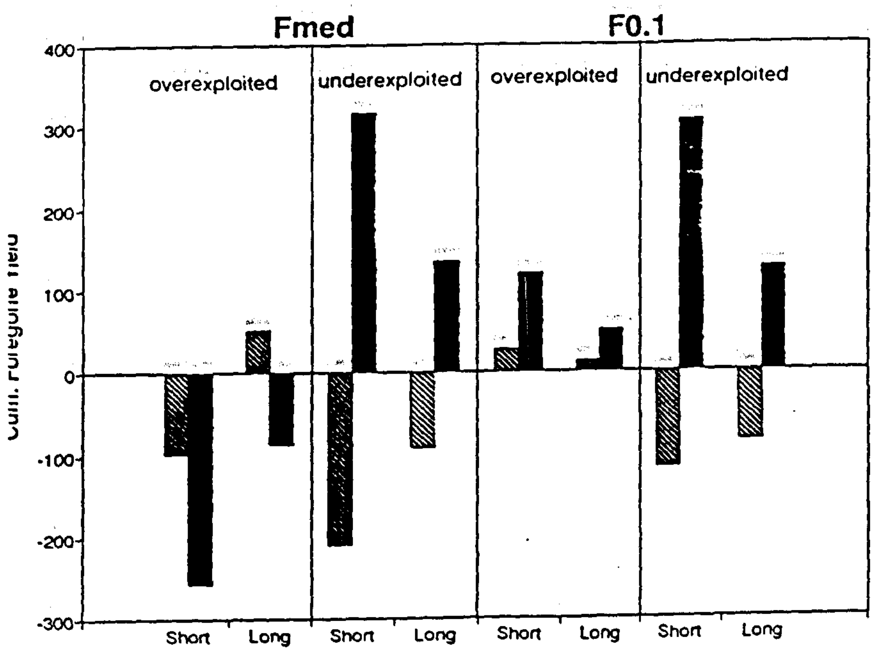
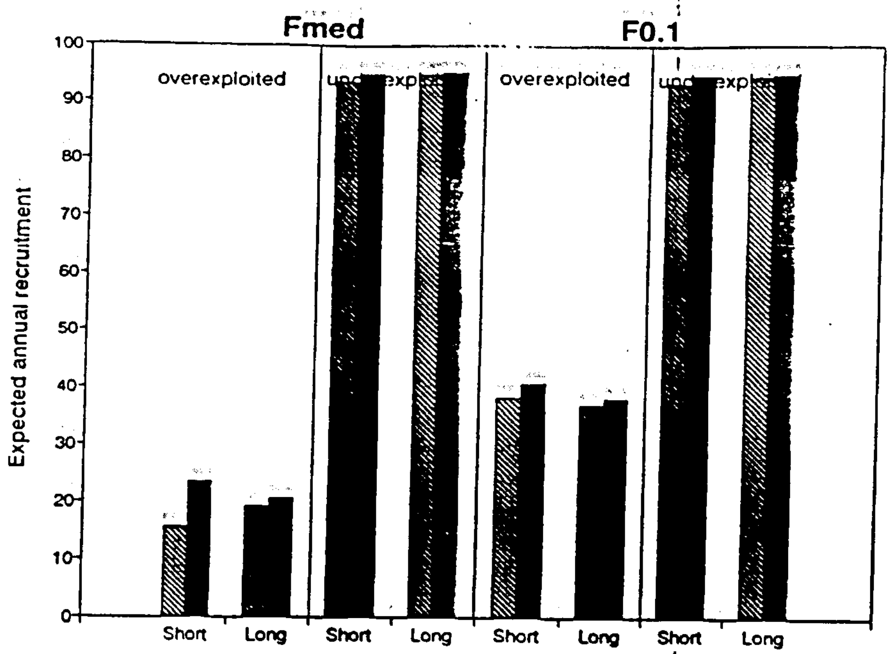


Figure 3: A. The coefficient of variation in annual yield across years and realizations. B. cumulative yield foregone compared to the case where th management strategy was implemented without error.

A.



B.

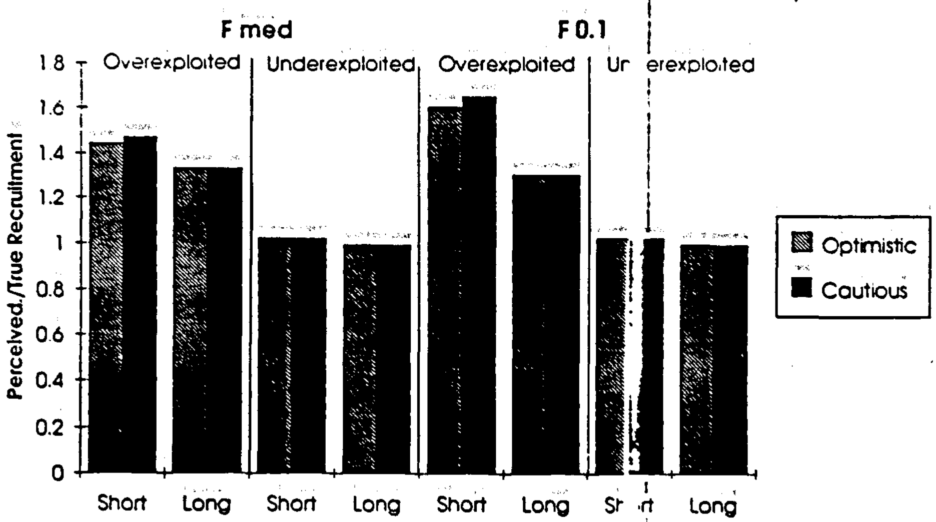
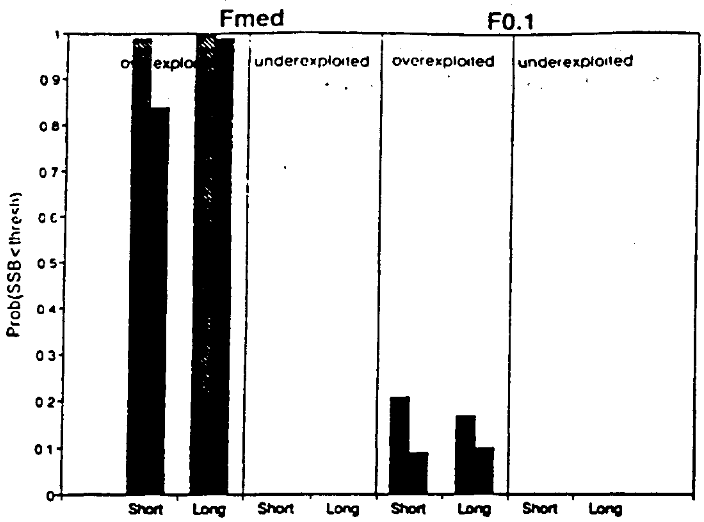
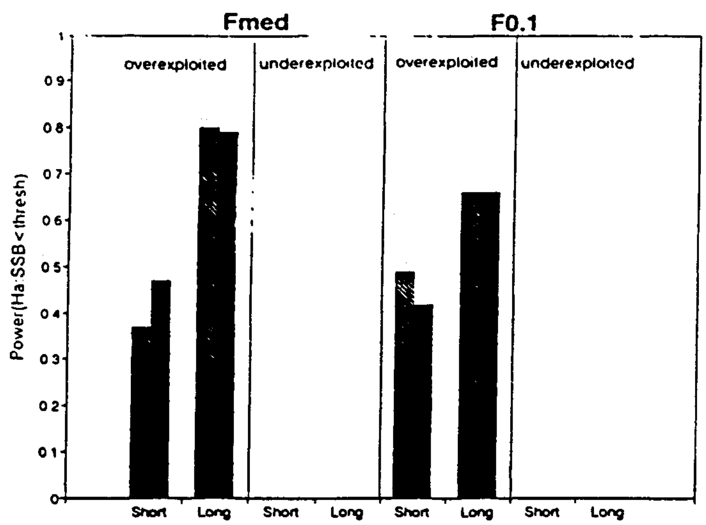


Figure 4: Expected annual recruitment averaged across years and realizations.
B. Ratio of perceived to true recruitment averaged across years and realizations.

A.



B.



C.

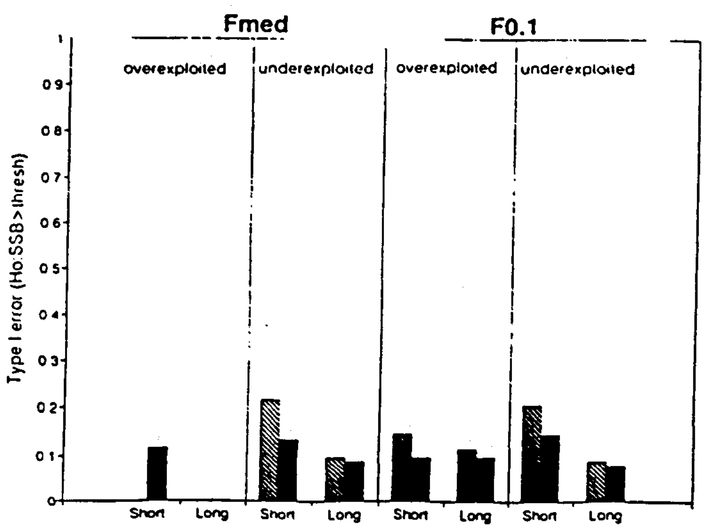


Figure 5: A. Probability that the true biomass was below the threshold 1/2 R max. B. Power of detecting that the biomass was below the threshold. C. Type I error for detecting biomass below