
Censored catch data in fisheries stock assessment

T.R. Hammond^{a*} and V.M. Trenkel^b

^a 9 Grove Street, PO Box 1012, Dartmouth, N.S., Canada, B2Y 3Z7

^b Département EMH, IFREMER, Rue de l'Île d'Yeu, BP 21105, 44311 Nantes, France

*: Correspondence to T. R. Hammond: tel: +1 902 426 3100 287; fax: +1 902 426 9654.

Abstract: Landings statistics can be lower than true catches because many fish are discarded or landed illegally. Since many discards do not survive, treating landings as true catches can lead to biased stock assessments. This paper proposes treating catch as censored by bounding it below by the landings, L , and above by cL (for scalar $c > 1$). We demonstrate the approach with a simulation study, using a Schaefer surplus production model. Parameters were estimated in a Bayesian framework with BUGS software using two sets of priors. Both the traditional true-catch method and a survey-and-effort method (which was landings free) performed worse on average than the censored approach, as measured by the Bayes risk associated with estimates of maximum sustainable yield (MSY) and of an index of depletion (X). Recursive partitioning (regression trees) was used to associate simulation parameters to best-performing methods, showing that higher commercial fish catchability favoured the censored method at estimating X . In conclusion, censored methods provide a means of dealing with discarding and misreporting that can outperform some traditional alternatives.

Keywords: censored data; discarding; misreporting; stock assessment

Introduction

While most fisheries stock assessments rely on reported landings, assessment methods differ in their treatment of these data. Methods derived from virtual population analysis (VPA), like extended survivor analysis (Shepherd, 1999), a standard ICES tool, assume that catches are known without error. For many stocks, however, this assumption doesn't hold, primarily because of discarding and misreporting. Observer studies suggest that the number of fish discarded is sometimes greater than the number landed for North East Atlantic megrim, whiting and Norway lobster (Rochet, et al., 2002). While the importance of misreported (and unreported) landings is more difficult to assess, anecdotal reports suggest these can be substantial. Thus, some models admit random errors in landings reports by assuming various error distributions centred on the observed landings (e.g. Kimura, 1990; Patterson, 1998). Still other models, e.g. Cook (1997), have dispensed with the landings data altogether, preferring to base results on survey data alone. This paper proposes a new approach: treating catch as censored data.

We don't mean the word 'censored' in the more familiar sense, as applied to material deemed inappropriate, but in the sense in which it is used in survival analysis. In the clinical trials typical of this field, a patient might still be alive at the end of the study period and thus his actual post-treatment survival time remains unknown (though, of course, it must be greater than the time from treatment to study's end). In such a case, we say the patient's survival time is censored. Thus, in calling the catch censored, we mean that we have observed bounds for the catch, but not the value itself.

In a previous comparison of catch-modelling approaches using simulated data, Patterson (1998) found that, in terms of mean square errors, a random catch error model provided better

results than ADAPT (a VPA-derived model) for a range of underreporting levels and survey index variances. It also outperformed a survey-only method. This paper also uses simulation to compare approaches, but deals with surplus production models rather than age-structured ones.

The most comprehensive study to date comparing the performance of various stock assessment models using simulated data was published by the National Research Council (NRC, 1998). One of the scenarios tested was a 30% underreporting of catches for which the surplus production model performed comparably well, in terms of the relative error of estimated exploitable biomass at the end of the study period (year 30). Curiously, the relative error was lower when only commercial catch data was used (6 %) compared to both commercial and survey data (24%). TAC estimates were completely wrong (366-179%), however, and other assessment methods performed much better.

Most other studies of the performance of surplus production models allowing for errors in commercial catches treat those errors as normal or lognormal. A notable exception is Koonce and Shuter, (1987) who assumed fixed underreporting of 20% but did not compare scenarios with and without underreporting.

Fisheries science has tried to deal with the discarding problem by placing observers on fishing vessels. By assuming that the fishermen behave no differently when they are not being observed, the observed discards are then extrapolated to fishing trips taken without observers and even to trips taken on vessels that never had any observers at all. The discards are said to be 'raised to the fleet' by means of such extrapolation. It is common practice to add the raised discards to the landings and to treat the sum as the true catch (or more recently as the 'observed' catch) for the fishery. Moreover, discards in years without observer data are

estimated using discards-to-landings ratios observed in different years (e.g., Anonymous, 2001).

This paper does not evaluate the merits of the above procedures. Instead, it focuses on a situation where the stock assessment working group has in hand a statistic, $L(t)$, which can safely be considered a lower bound for the true catch in year t . We will call $L(t)$ ‘the landings’ because we tend to imagine a situation where, despite considerable discarding, there is a total lack of observer data, but $L(t)$ could also be interpreted in other ways. It might be, for example, that there are so many anecdotal reports of illegal (black) landings that even the sum of observed landings and raised discards (now $L(t)$) is a lower bound for true catch. Alternatively, the working group might suspect that observer data are not representative of the whole fleet, so it is not prepared to raise the discards. In that case, $L(t)$ might simply represent the sum of observed discards (unraised) and landings. Thus, there are several possible scenarios for the situation we consider.

In addition to the lower bound, $L(t)$, we also assumed that the working group would be able to place a reasonable upper bound on the catches, which has important repercussions for the usefulness of the approach. It is convenient if this upper bound is expressed as a multiple of $L(t)$. If the working group were prepared to accept the premise that fishermen were not discarding more fish than they were landing, for example, then a reasonable upper bound would be $2 L(t)$. This is the value used throughout this paper.

Surplus production models are commonly used for stock assessments in situations where landings and effort data are available, but detailed age (or length) structure data are not (Hilborn and Walters, 1992). For this class of models, at least, we expected that treating landings as equal to catches, when catches are actually underreported, should adversely affect

stock biomass estimates. Thus, we used simulated data to evaluate the performance of three assessment methods: survey-and-effort, true-catch and censored catch. The survey-and-effort method tried to estimate catches with fishing effort data (assumed to be known without error), but relied primarily on surveys to estimate abundance. The true-catch method supplemented the survey index by treating recorded landings as true catch, and the last method also used the survey index but treated catch as censored data. All estimation methods were implemented using Bayesian techniques.

Bayesian analysis has become frequent in fisheries stock assessment (Hilborn and Liermann, 1998; McAllister and Ianelli, 1997; Meyer and Millar, 1999b; Punt and Hilborn, 1997; Walters and Ludwig, 1994) because of its clear, probabilistic portrayal of uncertainty and its natural extension to decision analysis (Punt and Hilborn, 1997). Bayesian probability, however, may be subjective in nature, and usually does not have the common interpretation that involves the frequency of an event. Various methods for obtaining posterior parameter distributions exist, one of which is Markov Chain Monte Carlo (MCMC). MCMC can be implemented with relatively little programming effort using freely available software called BUGS (www.mrc-bsu.cam.ac.uk/bugs/). (Meyer and Millar, 1999b) provide an application of BUGS to a surplus production model, and we used both their model and code as a template. The BUGS environment allowed us to change a standard Schaefer model into a censored data version by modifying only a few lines of code.

While BUGS facilitates the implementation of censored data models, and the Bayesian approach may aid in interpreting their results, there is nothing inherently Bayesian about censored data. In fact, many models in survival analysis do not follow this paradigm. Nor are the possible applications of this idea in fisheries restricted to surplus production models. A key point of this paper is that any statistical assessment model can be modified to treat catch

as censored data. Whether a given model should be so modified, however, should depend on the degree to which landings underestimate the true impact of fishing.

Methods

Censored Data

For a given random variable Y , with density f and distribution function F , a standard data set might consist of two realizations, $y_1 = 3$ and $y_2 = 5$, of Y . A set of censored data, on the other hand, might have observations of the form $y_1 > 3$ and $y_2 < 5$. Whereas the likelihood (l) for the standard data set would be computed using the density function ($l = f(3)f(5)$), in this simple example the censored likelihood would be computed from the distribution function ($l = (1-F(3))F(5)$). Maximum likelihood (or Bayesian) estimation of the parameters of f (and of F) could proceed much as before, though the likelihood function would be less informative. Of course, if there were other random variables correlated with Y , then the situation becomes more complicated, as the censored data for Y would be informative about them. In the frequentist context, special algorithms are used to handle censored data problems. Even the most common of these, the Estimation Maximization (EM) algorithm (Tanner, 1996; McLachlan and Krishnan, 1997), is very rarely applied in fisheries. In the Bayesian context, however, the same algorithms that are typically used in fisheries stock assessment, namely MCMC (Meyer and Milar, 1999a) or SIR (McAllister and Ianelli, 1997), can readily be modified to handle censored data, as demonstrated here for MCMC. This paper uses censored observations of the form $L(t) < C(t) < 2 L(t)$, where $C(t)$ is catch in year t and $L(t)$ represents reported landings (but see above).

Experimental Design

The simulation experiment in this paper was designed to compare the performance of three different stock assessment approaches, all based on a discrete variant of the Schaefer surplus production model (Meyer & Millar, 1999):

$$(1) \quad B'(t+1) = B'(t) (1+r) - rB'(t)^2 - \frac{C(t)}{K}.$$

In Equation (1), $B'(t)$ is biomass ($B(t)$) divided by carrying capacity (K) in year t , r is the intrinsic population growth rate, and $C(t)$ is the catch. The three approaches differed in their treatment of catch. All simulated stocks followed Schaefer model population dynamics, a simplification that we justify by our focus on catch rather than on the effects of model process error. We did, however, add a degree of hyper-depletion or hyper-compensation (c.f. Hilborn and Walters, 1992) to the simulated data, so that catch-per-unit-effort would not be proportional to biomass. As a result, no assessment model matched the simulation model perfectly. The data available to the stock assessment models were annual landings, fishing effort, and scientific survey catch rates. All simulated data sets contained ten years worth of data.

We took a decision theoretic approach to the comparison of estimation approaches, grading performance in terms of a number called the Bayes risk (e.g. Casella and Berger, 1990). In fact, the paper's main objective is to determine which approach has the lowest such risk. But this simple comparison is complicated by the fact that the Bayes risk can be computed for any model parameter. Moreover, the computation can be made with a variety of estimation loss functions. To constrain the computations, therefore, we chose just two parameters that we considered particularly important to stock assessment: maximum sustainable yield ($MSY =$

$rK/4$) and depletion (X), the latter defined as the final-year biomass divided by the biomass that yields MSY (namely $K/2$, which implies that $X = 2B'(10)$, year 10 being the final year).

We also chose two loss functions: squared-error loss (λ_{SE}) and absolute-error loss (λ_{AE}).

These decisions gave us four Bayes risk measures to consider for each estimation method.

Loss functions are designed to score estimation performance by comparing an estimation result ($\delta_p(\mathbf{x})$, for a particular parameter p given data vector \mathbf{x}) with the ‘true’ value of that parameter (which, for the sake of generality, we take to be a scalar valued function of the ‘true’ parameter vector ξ - denoted $p(\xi)$). In other words, loss is just a function of ξ and \mathbf{x} that estimators should try to minimize. As we consider Bayesian estimation, we take $\delta_p(\mathbf{x})$ to be the marginal posterior density for p given the data \mathbf{x} - denoted here by $\pi_p(\rho | \mathbf{x})$ (for dummy variable ρ). In this notation, our two estimation loss functions are defined as follows:

$$(2) \quad \lambda_{SE}^p(\xi, \mathbf{x}) = \int (p(\xi) - \rho)^2 \pi_p(\rho | \mathbf{x}) d\rho$$

$$(3) \quad \lambda_{AE}^p(\xi, \mathbf{x}) = \left| \int (p(\xi) - \rho) \pi_p(\rho | \mathbf{x}) d\rho \right|$$

Squared-error loss bears a strong resemblance to the classical mean squared error, while absolute-error loss is not unlike the absolute value of the estimation bias.

The loss functions are difficult to evaluate analytically, but they are very well approximated by the results of MCMC. To approximate these equations using n sample values p_1, p_2, \dots, p_n from the posterior distribution $\pi_p(\rho | \mathbf{x})$ given by MCMC, we used the following:

$$(4) \quad \lambda_{SE}^p(\xi, \mathbf{x}) \approx \frac{1}{n} \sum_{i=1}^n (p_i - p(\xi))^2 \quad \text{or}$$

$$(5) \quad \lambda_{AE}^p(\xi, \mathbf{x}) \approx \frac{1}{n} \left| \sum_{i=1}^n (p_i - p(\xi)) \right|$$

Now the basic idea of the Bayes risk is to obtain an average value of the estimation loss by integrating over Ξ (the parameter space of possible values for ξ) and over \mathcal{X} (the sample space for random vectors \mathbf{x}). For squared error loss, the formula for the Bayes risk ? (after Casella and Berger, 1990) is as follows:

$$(6) \quad B_{SE}^p = \int_{\Xi} \int_{\mathcal{X}} \lambda_{SE}^p(\xi, \mathbf{x}) f(\mathbf{x} | \xi) \pi(\xi) d\mathbf{x} d\xi ,$$

where $f(\mathbf{x}|\xi)$ is the likelihood function and $\pi(\xi)$ is the prior. Naturally, B_{AE}^p , the Bayes risk for parameter p under absolute-error loss, would be defined analogously. Notice that the Bayes risk places the highest weight on the most *a-priori* probable values of ξ and \mathbf{x} .

In practice, Equation 6 presents us with a very daunting integration problem. We certainly do not recommend that analytic solutions be attempted. Instead, we propose to approximate by Monte Carlo integration, just as in equations (4) and (5). If we first notice that the term $f(\mathbf{x}|\xi) \pi(\xi)$ in equation (6) is just the joint distribution of \mathbf{x} and ξ , then it will come as no surprise that, given m vectors $(\mathbf{x}_1, \xi_1), (\mathbf{x}_2, \xi_2), \dots, (\mathbf{x}_m, \xi_m)$ drawn randomly from that joint distribution, the Bayes risk can be approximated as follows:

$$(7) \quad B_{SE}^p \approx \frac{1}{m} \sum_{i=1}^m \lambda_{SE}^p(\mathbf{x}_i, \xi_i)$$

Within equation (7), the λ_{SE}^p term can also be approximated by equation (4). And, of course, an analogous formula would apply to B_{AE}^p .

Our approximation to the Bayes risk affected the design of our simulation experiment, and as a result it differs considerably from the usual frequentist approach. A frequentist simulation study would normally select a set of Schaefer model parameters and simulate many data sets with them. Such a study would consider several scenarios with different model parameters. In this study, on the other hand, the two scenarios were defined by different prior distributions. For each simulation within a scenario, we first drew a set of model parameters (ξ) from the priors and then constructed a data set (\mathbf{x}) consisting of survey, landings or effort data by drawing randomly from $f(\mathbf{x}|\xi)$. In other words, both parameters and data varied over the simulations. This simulation scheme provides the joint vectors (\mathbf{x}_1, ξ_1) , (\mathbf{x}_2, ξ_2) , ..., (\mathbf{x}_m, ξ_m) in just the form needed to approximate the Bayes risk by equation (7). We did $m = 40$ simulations for each scenario.

It is natural to wonder at this point whether the 40 simulations we did is ‘enough’. To address this issue, let us first consider what ‘enough’ should mean in this context. We contend that we do not need to evaluate the Bayes risk to any particular degree of precision that might be chosen *a-priori*. Instead, we need only determine, with a high degree of probability, which assessment method had the lowest risk. This implies that the number of simulations (m) required is closely linked to the actual performance difference between methods: the more they differ in performance, the fewer the simulations required. So, to determine whether the observed difference between the top two performing methods was ‘significant’, we ran a paired, Wilcoxon rank sum test (two-sided, level $\alpha = .0001$) comparing the 40 calculations of λ_{AE}^x (with apologies to the Reverend Bayes). We repeated the test

with λ_{SE}^{MSY} , λ_{SE}^X , and λ_{AE}^{MSY} . If all these tests are significant at that low level, then 40 is enough.

The Bayes risk is effective at evaluating the overall performance of the various methods, but hides the effects of different parameter values. Therefore, we used regression trees to investigate the link between simulation model parameters and which method was optimal for them. Regression trees are an exploratory tool suitable for modelling categorical response variables (here which model performed best in terms of λ_{SE}^p) using categorical or continuous explanatory variables (Chambers and Hastie, 1992). Individuals (here simulated data sets) are recursively split into two groups using the explanatory variables (here model parameters plus average ratio of catch to landings). At each node the split leading to the greatest difference in the response variable is selected, and these splits are represented as branches of a tree. The leaf nodes of this tree are clusters of similar individuals. By looking at the values of the variables characterising each branch, a description of the final cluster is obtained. It is possible to set a minimum group size for clusters (we used 8) in order to avoid clusters with only few individuals.

Simulation

We started each simulation by generating a new ten-year time series of fishing effort, $E(t)$. These time series were constructed by simulating three numbers from a Uniform distribution on the interval from 0.2 to 1.2. These three numbers represented the effort in years one, six and ten. The other effort values were obtained by linear interpolation. In other words, the effort time series was not affected by any trends in the size of the fished population. We justify this simplification by the fact that, in reality, the effect of biomass trends on effort can be quite weak, given the many other factors at play (fuel prices, other stocks, other jobs, government subsidies, etc.).

The next step was to draw a set of Schaefer assessment model parameters from the prior distributions. We used two scenarios representing different levels of prior information about the stock. These two scenarios are described in Table 1, which also gives a description of each model parameter. The principal difference between the scenarios is that, in the second, both the survey data and the priors were considerably less informative. These priors were chosen so as to simulate population trajectories that we considered realistic.

Table 1. The prior distributions used in both evaluation scenarios. In these priors, the Gamma distribution family is parameterised so as to have mean equal to the first parameter divided by the second. The Lognormal distribution is parameterised by the mean and variance of the associated Normal distribution, and the Beta(α, β) distribution is parameterised to have mean equal to $\alpha/(\alpha+\beta)$.

Parameter	Description	Scenario 1	Scenario 2	BUGS Alias
r	Intrinsic growth rate	logN(-1.4,0.25), mean=0.28, CV=0.53	logN(-2.8, 1/3), mean=0.07, CV=0.63	r
K	Carrying capacity	logN(log(700), 0.25) , mean= 793, CV=0.53	logN(log(1000), 0.25), mean=1133, CV=0.53	K
σ^{-2}	Process error precision	Gamma(100, 0.05), mean=2000, CV=0.10	Gamma(120, 0.05), mean=2400, CV=.09	isigma2
θ^{-2}	Catch model precision	Gamma(70, 1), mean=70, CV= 0.12	Gamma(50, 1) mean=50, CV=0.14	itheta2
τ^{-2}	Survey precision	Gamma(44, 2), mean=22, CV= 0.15	Gamma(25, 2), mean=12.5, CV=0.2	itau2
q_c	Fleet catchability	Beta(5.5,44.5), mean=0.11, CV=0.40	Beta(5,44) mean=0.10, CV=0.42	qcomm
q_s	Survey catchability	Beta(4,8), mean=1/3, CV=0.39	Beta(2,4), mean=1/3, CV=0.53	q

The simulations created fish stocks that were already depleted below the carrying capacity, K , in year one. This was done by drawing the initial proportion of K , denoted P_0 , from a beta

distribution:

$$(8) \quad P_0 \sim \text{Beta}(20,6).$$

We assumed the biomass in the first year was normally distributed, with process error variance σ^2 :

$$(9) \quad B'(1) \sim N(P_0, \sigma^2),$$

As mentioned above, the simulation model added a bit of hyper-depletion or hyper-compensation to the relationship between catch rates and biomass. This was achieved using a uniformly distributed random variable called δ ,

$$(10) \quad \delta \sim U(0.7, 1.3)$$

This done, we created the catch and biomass time series:

$$(11) \quad C(t) \sim \log N\left(\log(q_c E(t) B'(t)^\delta K), \theta^2\right)$$

$$(12) \quad B'(t+1) \sim N\left((r+1)B'(t) - rB'(t)^2 - \frac{C(t)}{K}, \sigma^2\right)$$

The lognormal distribution is parameterised in this paper by the mean and variance of the associated normal distribution. In equation (6), we used the same process error variance from equation (3).

We generated a mean landings rate (p_L) for our simulated fishery:

$$(13) \quad p_L \sim U(0.5, 0.9)$$

The actual landings rate in year t , namely $p_L(t)$, was taken to be uniformly distributed in the

interval $[p_L - 0.1, p_L + 0.1]$ Then the landings were obtained as follows:

$$(14) \quad L(t) = p_L(t) C(t)$$

Equations (13) and (14) imply that the simulated catches in a given year could be up to 2.5 times greater than the landings.

The survey catch rates, denoted $I(t)$ in year t , were generated next. We assumed a lognormal distribution for the survey index:

$$(15) \quad I(t) \sim \log N(\log(q_s KB'(t)), \tau^2),$$

where q_s is the constant survey catchability.

Estimation

The three estimation models, survey-and-effort, true-catch and censored catch, closely resemble the simulation model in that they share all its equations except (10), (11), (13) and (14). Equation (10), above, defines parameter δ , which was used to create a non-linear relationship between commercial catch rates and biomass in the simulated datasets. This parameter was not used at all in estimation. For the true-catch method, which we consider traditional because it is like the one in Meyer and Millar, (1999a), catches were assumed equal to landings, $C(t) = L(t)$, so there was no need for equations (11), (13) and (14). For the survey-and-effort and censored catch methods, catches were predicted from the product $q_c E(t) B(t)$, where q_c is the catchability coefficient for commercial fishing and $E(t)$ is the corresponding effort in year t . This is equivalent to assuming that catch rates are proportional to biomass. Thus equations (13) and (14) were eliminated and (11) was replaced with the following (wherein the δ has simply been deleted):

$$(16) \quad C(t) \sim \log\text{N}(\log(q_c E(t)B'(t)K), \theta^2).$$

In the censored method, Equation (16) was interpreted in the light of the censored catch data, which took the form $L(t) < C(t) < 2L(t)$, for every year t . The survey-and-effort model made no use of the landings data.

The likelihood function for the data involves a factor for the survey index:

$$(17) \quad \prod_{t=1}^{10} \text{dlogN}(I(t), \log(q_s B'(t)K), \tau^2),$$

and another for the biomass

$$(18) \quad \text{dN}(B'(1), P_0, \sigma^2) \prod_{t=2}^{10} \text{dN}(B'(t+1), (r+1)B'(t) - rB'(t)^2 - \frac{C(t)}{K}, \sigma^2),$$

where $\text{dlogN}(a, b, c)$ denotes the $\log\text{N}(b, c)$ density function evaluated at a , and $\text{dN}(a, b, c)$ denotes the $\text{N}(b, c)$ density function evaluated at a . In the survey-and-effort and censored methods, the likelihood also includes a factor for the catch:

$$(19) \quad \prod_{t=1}^{10} \text{dlogN}(C(t), \log(q_c E(t)B'(t)K), \theta^2),$$

but not in the true-catch method, where $C(t) = L(t)$,

Fortunately, within the BUGS programming environment, the full joint likelihood need not be specified, as the program can deduce it from the conditional distributions and from the conditional independence relationships implied by their specification. Appendix 1 provides the BUGS code that we used to implement the true-catch method. This appendix also indicates the modifications required to adapt this code to the survey-and-effort and censored-

catch models.

For each simulation we ran two MCMC chains for one million iterations, sampling every 40th iteration, after a burn-in of 500 000 iterations. These two chains were started from different, over-dispersed starting locations, as specified in Appendix 1. Running two chains allowed us to compute the relative difference between the marginal posterior quantiles for each parameter. This provides an intuitive measure of MCMC-induced error, more reliable at examining many chains than statistical convergence tests. All results in this paper were based on the combined samples in both chains.

Results

In order to assess the MCMC-induced error, we took our two MCMC chains and computed the difference between estimates of the first, second and third quartiles of every parameter, and divided that difference by the mean parameter value (this last an average over both chains). The maximum absolute value of this relative difference, over all parameters and over the three quartiles, provides an index of the relative difference between the chains. We looked at the average of that index over all 40 simulations. For the censored method (in scenario 1) the average was 0.89%; for the survey-and-effort method it was 1.87% and for the true catch it was 1.09%. The worst case for this relative difference index (over the simulations) was 2.37% for censored, 4.72% for survey-and-effort and 2.93% for true catch. This analysis revealed that late-year catches mixed most slowly in the MCMC chains. The MCMC-induced error was small compared to the relative difference between results across estimation methods (using the same MCMC start values- where possible), which averaged over 50% (over the parameters and quantiles in scenario 1). Thus, MCMC performance was deemed acceptable for the purpose of comparing methods on simulated data.

The Bayes risk for depletion (X) under squared-error loss (B_{SE}^X) was smallest for the censored catch method in scenario 1, at 0.077. For the survey-and-effort method this statistic was 0.098 and for the true catch method it was 0.108. For B_{SE}^{MSY} the results were 2279 for the censored method, 3098 for the survey-and-effort and 2728 for the true catch in the same scenario. With absolute-error loss in scenario 1, the results were 0.162 for the censored, 0.178 for the survey-and-effort, and 0.214 for the true catch methods for B_{AE}^X . For B_{AE}^{MSY} the results were 20.25 for the censored method, 25.75 for the survey-and-effort and 23.35 for the true catch.

Figure 1 gives a graphical representation of these estimation results. The first row of this figure provides marginal posterior results from estimating the depletion index X , while the second row gives results for MSY . These marginal distributions have been shifted left by the true parameter value (i.e. the one generated in simulation), so the results should be clustered around zero, if all is going well. All methods are a bit over-optimistic about the stock, but the true catch method is particularly so.

Results were much the same in scenario 2: B_{SE}^X was 0.109 for censored, 0.140 for survey-and-effort and 0.201 for true catch. B_{SE}^{MSY} was 1675 for the censored, 3441 for the survey-and-effort, and 2401 for the true catch. In terms of absolute-error loss, the B_{AE}^X results were 0.226 for censored, 0.245 for survey-and-effort and 0.335 for true catch. For B_{AE}^{MSY} , the censored model had 26.1, the survey-and-effort had 35.7 and the true catch had 27.0. The overall decline in performance in this scenario reflects the less informative survey data. As before, Figure 2 gives a graphical representation of the estimation results. There is a more pronounced trend in this scenario towards overly optimistic estimates of MSY , especially in

the true catch method.

Figure 1. These are box and whisker plots showing the marginal posterior for the depletion index X (top) and for MSY (bottom) in each simulation of scenario 1. Estimates are centred on the 'true' (i.e. the simulated) values. (a) censored model, (b) survey-and-effort model, (c) true-catch model. Each box shows the first, second and third quartiles, while the whiskers delimit 95% confidence intervals.

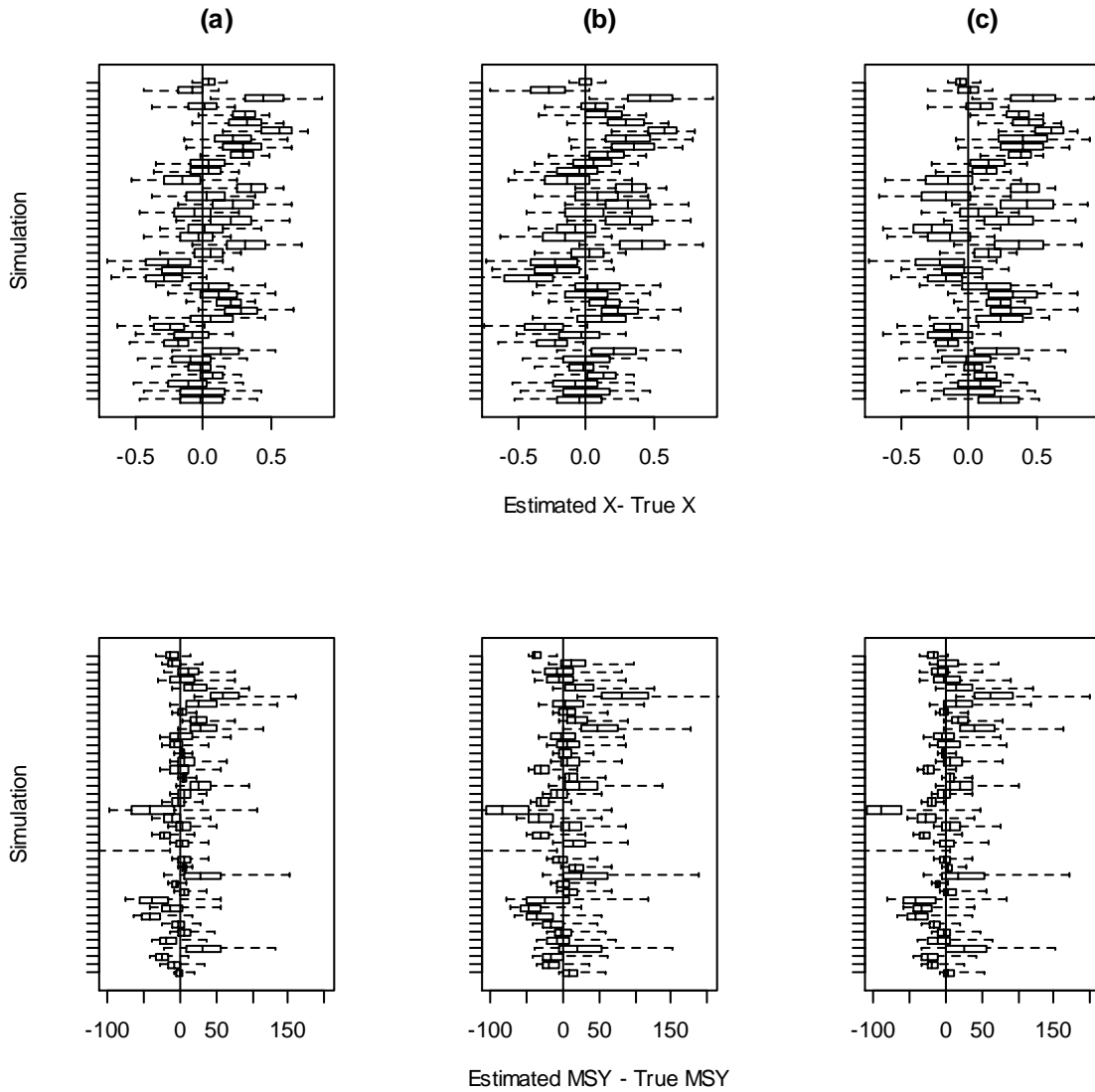
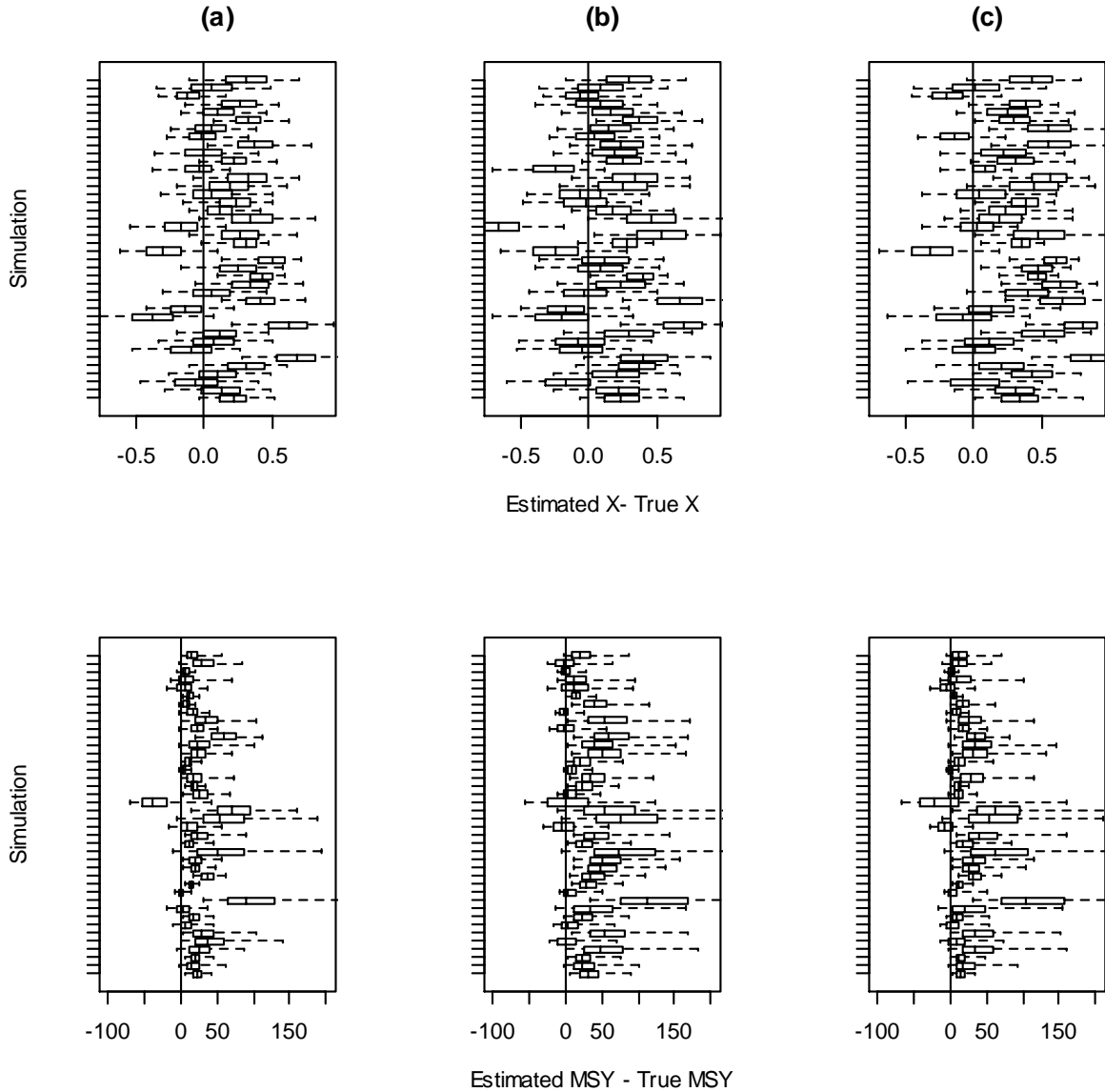


Figure 2. These are box and whisker plots showing the marginal posterior for the depletion index X (top) and for MSY (bottom) in each simulation of scenario 2. Estimates are centred on the ‘true’ (i.e. the simulated) values. (a) censored model, (b) survey-and-effort model, (c) true-catch model. Each box shows the first, second and third quartiles, while the whiskers delimit 95% confidence intervals.



All the Wilcoxon rank sum tests comparing values of λ_{SE}^{MSY} , λ_{SE}^X , λ_{AE}^{MSY} and λ_{AE}^X were significant at level $\alpha = .000001$, showing that the $m = 40$ simulations used in each scenario were sufficient to discern that the censored method’s superior performance is unlikely to have occurred by chance.

Thus, the overall performance of the censored approach was superior, but it was decidedly not optimal in several situations. With recursive partitioning we attempted to characterize the situations in which each method performed best. For this purpose, squared-error loss was chosen as the optimality criterion. For each simulation (40 for each scenario) the dependent variable was coded by the method with the minimum λ_{SE}^p (censored, survey-and-effort or true catch). The resulting trees are shown in Figures 3 through 6, for MSY and X in scenarios 1 and 2. The dominating method with the smallest λ_{SE}^{MSY} is marked at the end of each branch. For example, branch 1 in Figure 3 consists of 20 simulations for which the censored method generally performed best. Some misclassification can happen, hence the percentage of simulations classified correctly is also given in the figures. Looking at these Figures for patterns that are consistent between scenarios, we find that higher fish catchability (q_c) values and lower precision for the survey index (τ^{-2} in scenario 1) favoured the censored method at estimating the depletion index X compared to the other two methods. Higher ratios of catch to landings (ratio), i.e. less discarding, favoured it at estimating MSY in scenario 2, but not in scenario 1. Note that the survey-and-effort method was relatively effective at estimating depletion, but poor at estimating MSY .

Figure 3. Regression tree for scenario 1 linking simulation parameters to the best performing method, as measured by the mean squared error in MSY . Each branch of the tree gives a particular parameter and a branching rule for that parameter. The branching rule is specified by $<>$ or by $><$, which indicates which way to branch. So, for ' $r >< 0.235$ ' you branch left when $r > 0.235$ and right when it isn't. The leaves of the tree are numbered, and they indicate the predicted best performing method using the following code: cen =censored catch, tc=true catch, s&e=survey-and-effort. See Table 1 for a mapping of parameter aliases to the symbols used in the Methods section.

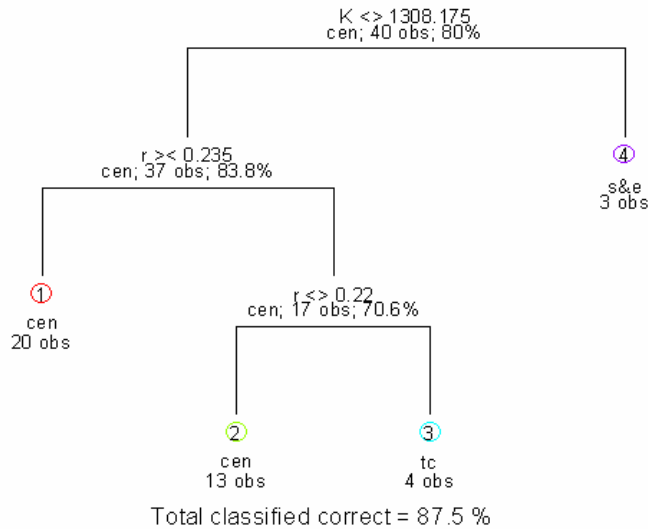


Figure 4. Regression tree for scenario 1 linking simulation parameters to the best performing method, as measured by squared-error loss estimating depletion (λ_{SE}^X). Each branch of the tree gives a particular parameter and a branching rule for that parameter. The branching rule is specified by $<>$ or by $><$, which indicates which way to branch. So, for ' $ratio <> 1.79$ ' you branch left when $ratio < 1.79$ and right when it isn't. The leaves of the tree are numbered, and they indicate the predicted best performing method using the following code: cen =censored catch, tc=true catch, s&e=survey and effort. Parameter 'ratio' is the average ratio of catch to landings.

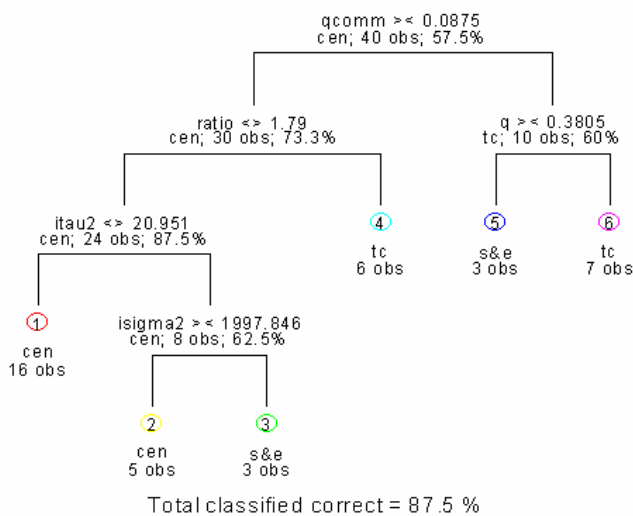


Figure 5. Regression tree for scenario 2 linking simulation parameters to the best performing method, as measured by λ_{SE}^{MSY} , the squared-error loss for *MSY*. Each branch of the tree gives a particular parameter and a branching rule for that parameter. The leaves of the tree indicate the predicted best performing method using the following code: cen=censored catch, tc=true catch, s&e=survey and effort. Parameter 'ratio' is the average ratio of catch to landings.

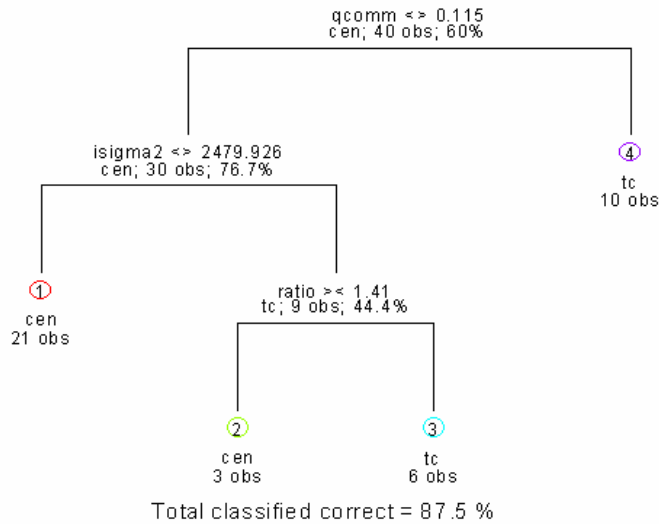
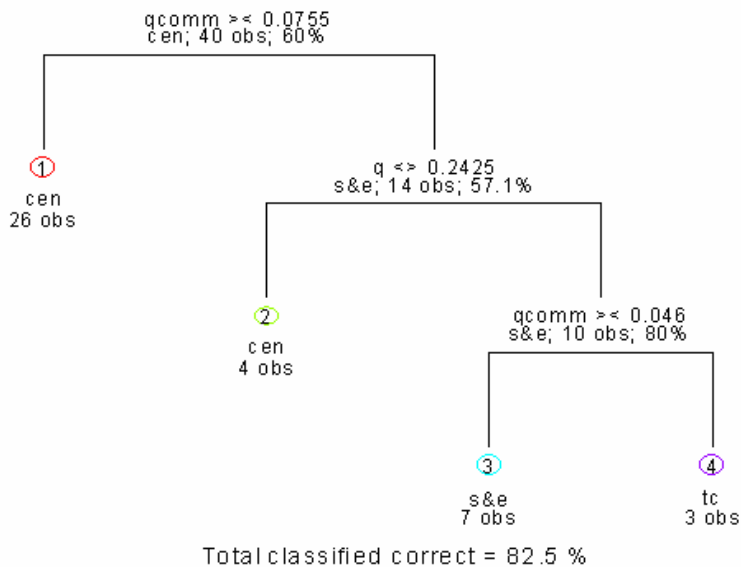


Figure 6. Regression tree for scenario 2 linking simulation parameters to the best performing method, as measured by λ_{SE}^X the squared-error loss for parameter *X* in scenario 2. The leaves of the tree indicate the predicted best performing method using the following code: cen=censored catch, tc=true catch, s&e=survey and effort. See Table 1 for a documentation of the parameters in this figure.



Discussion

This paper introduces the idea of treating catch as censored data in stock assessment models. Though censored data have been used for decades in survival analysis, the concept seems to be new to fisheries. In this paper it was shown that censored data models can lead to improved performance (on average) relative to the traditional true-catch and survey-and-effort methods on simulated data sets, both at estimating depletion and at estimating *MSY*. This was true regardless of whether performance was measured by squared-error or absolute-error loss, and it did not appear to be sensitive to the prior distributions. The censored method was not, however, always the best performer.

Although we explored which parameter values were associated with good performance of the censored method, few patterns emerged that were consistent across scenarios. The only such pattern that we uncovered suggested that the censored method was favoured by high catchability values when estimating depletion. We had expected that the relative performance of the censored and true catch methods would depend of the average ratio of catch to landings, with higher ratios favouring the censored approach. This turned out to be the case in estimating *MSY* in scenario 2, but not in scenario 1. Without clear indications of parameter values that give the censored method special difficulty, we decided that its superior average performance might be expected in general.

Fisheries science has increasingly relied on scientific surveys in stock assessment because of the difficulties with commercial fishery data. Some recommend dispensing with landings data altogether in certain situations (e.g. Cook, 1997), preferring landings-free methods. We showed here that our landings-free method, the survey-and-effort one, was relatively effective at estimating depletion, but not at estimating *MSY*. Nonetheless, our results suggest that

landings-free methods throw away too much information as the survey-and-effort method was beaten by the censored approach even at estimating depletion.

The censored catch idea can only prove useful, however, if real situations arise where catches are surely greater than landings. When catches are greater than landings, it is most often because fishermen discard many of the fish they catch at sea. Fish that are thrown back in the ocean after being hauled on board ship by commercial fishing gear are usually injured, if not dead. In many fisheries worldwide, it is not unusual for more fish to be discarded than landed (e.g. Alverson, et al. 1994, Stratoudakis, et al. 1999, Rochet et al., 2002). Though fishermen may regret the waste involved, market forces and fisheries regulations interact to compel them. Indeed, the use of landings quotas in mixed fisheries creates an incentive to discard fish because it is difficult to avoid catching a mixture of species (Pascoe, 1997). A vessel that has quota for cod and other stocks, for example, will not usually stop fishing when she fills her cod quota, but will keep going until she fills all quotas, discarding as many cod as necessary in the process. To ignore discarding in stock assessment, therefore, is usually to underestimate the true impact of fishing.

Fishing gears also kill many fish that they fail to capture. Though fisheries science has a long tradition of using trawl mesh-size regulations to allow unwanted small fish to slip through nets, video footage of gear performance suggests that many of the fish escaping capture in this way are still injured or killed, e.g. (Suuronen, et al., 1996). Many gear types also lose fish in the process of hauling them on board, drift nets being a prominent example. Ghost fishing, by which we mean lost fishing gear that continues to kill fish (Kaiser, et al., 1996), provides yet another mechanism by which landings may underestimate fishing impact.

In stock assessment it can be foolish to neglect the effects of lies and omissions. Many

fishermen feel compelled to land fish illegally, to misreport their catch, to mislabel it, or to misrepresent its origins. Poaching is a huge problem in fisheries around the world. For the most part, these ‘human failings’ lead to landings statistics that underestimate fishing impact, and reinforce the need for censored data models. But there are a few notable exceptions.

Despite compelling reasons for thinking that landings underestimate fishing impact, there are mechanisms that can lead to the very opposite. The herring fishery on the west coast of Scotland (area VIa) provides an interesting example. According to the stock assessment working group reports for this area, fishermen have been reporting fish that they caught in the North Sea (area IVa) in area VIa (to some degree also in areas IIIa and IIa), in order to take advantage of the less restrictive quota in VIa (ICES, 2002). This might mean that landings in VIa are greater than catches.

Such examples being exceptional, a lower bound for the catch could likely be found in most fisheries; however, the censored method also calls for an upper bound. This is a weak point because it is hard to imagine how such a bound could be directly observed; it would have to be set somewhat arbitrarily. Nonetheless, we feel that a stock-assessment working group should be able to set a level beyond which the catch would be incredible. We showed in this paper that, even if this upper bound is set too low (at 2 times landings rather than the 2.5 possible in simulated data), censored model performance can still be better than that of traditional approaches. Future use of censored data models may well depend on how this arbitrary aspect is viewed by working groups.

Acknowledgement

This work was partly funded by the European Commission under Framework V, project contract QLRT-1999-01609. It was also partly funded by Defence R&D Atlantic.

References

- Alverson, D.L., Freeberg, M.H., Murawski, S.A., and Pope, J.G. 1994. A global assessment of bycatch and discards. FAO Fisheries Technical Paper 339, 233 pp.
- Anonymous. 2001. Report of the working group on the assessment of Southern Shelf demersal stocks. ICES C.M. 2001 / ACFM: 5, 735 pp.
- Casella, G., and Berger, R. 1990. Statistical Inference. Wadsworth and Brooks, Belmont, 650 pp.
- Cook, R. M. 1997. Stock trends in six North Sea stocks as revealed by an analysis of research vessel. ICES Journal of Marine Science 54: 924-933.
- Hilborn, R., and Liermann, M. 1998. Standing on the shoulders of giants: learning from experience in fisheries. Reviews in fish biology and fisheries 8: 273-283.
- Hilborn, R., and Walters, C. J. 1992. Quantitative fisheries stock assessment: Choice, dynamics and uncertainty. Chapman and Hall, New York.
- ICES. 2002. Report of the herring assessment working group for the area South of 62° N, ICES CM 2002/ACFM/:12, 418 pp
- Kaiser, M. J., Bullimore, B., Newman, P., Lock, K., and Gilbert, S. 1996. Catches in 'ghost fishing' set nets. Marine Ecology Progress Series 145: 11-16.
- Kimura, D. K. 1990. Approaches to age-structured separable sequential population analysis. Canadian Journal of Fisheries and Aquatic Sciences 47: 2364-2374.
- Koonce, J. F. and Shuter, B. J. 1987. Influence of various sources of error and community

- interactions on quota management of fish stocks. *Canadian Journal of Fisheries and Aquatic Sciences* 44: 61-67.
- McAllister, M. K., and Ianelli, J. N. 1997. Bayesian stock assessment using catch-age data and the sampling - importance resampling algorithm. *Canadian Journal of Fisheries and Aquatic Sciences* 54: 284-300.
- McLachlan, G.J., and Krishnan, T. 1997. *The EM Algorithm and Extensions*. John Wiley and Sons. New York. 275 pp.
- Meyer, R., and Millar, R. 1999a. BUGS in Bayesian stock assessments. *Canadian Journal of Fisheries and Aquatic Sciences* 56: 1078-1086.
- Meyer, R. and Millar, R. B. 1999b. Bayesian stock assessment using a state space implementation of the delay difference model. *Canadian Journal of Fisheries and Aquatic Sciences* 56: 37-52.
- NRC (National Research Council). 1998. *Improving Fish Stock Assessments*. Washington, DC, National Academy Press, 177 pp.
- Pascoe, S. 1997. *Bycatch management and the economics of discarding*. FAO Fisheries Technical Paper no. 370. Rome: FAO.
- Patterson, K. 1998. Assessing fish stocks when catches are misreported: model, simulation tests, and application to cod haddock, and whiting in the ICES area. *ICES Journal of Marine Science* 55: 878-891.
- Pope, J. G., and Shepherd, J. G. 1982. A simple method for the consistent interpretation of catch-at-age data. *Journal du Conseil International pour l'Exploration de la Mer* 40: 176-

- Punt, A. E., and Hilborn, R. 1997. Fisheries stock assessment and decision analysis: the Bayesian approach. *Reviews in Fish Biology and Fisheries* 7: 35-63.
- Rochet, M. J., Péronnet, I., and Trenkel, V. M. 2002. An analysis of discards from the French trawler fleet in the Celtic Sea. *ICES Journal of Marine Sciences* 59: 538-552.
- 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 Sciences* 56: 584-591.
- Stratoudakis, Y., Fryer, R.J., Cook, R.M., and Pierce, G.J. 1999. Fish discarded from Scottish demersal vessels: estimators of total discards and annual estimates of targeted gadoids. *ICES Journal of Marine Science*, 56: 592-605.
- Suuronen, P., Perez-Comas, J., Lethonen, E., and Tschernijj, -. V. 1996. Size-related mortality of herring (*Clupea harengus* L.) escaping through a rigid sorting grid and trawl codend meshes. *ICES Journal of Marine Science* 53: 691-700.
- Tamsett, D., Janacek, G., Emberton, M., Lart, B., and Course, G. 1999. Onboard sampling for measuring discards in commercial fishing based on multilevel modelling of measurements in the Irish Sea from NW England and N Wales. *Fisheries Research*, 42: 117-126.
- Tanner, M.A. 1996. *Tools for Statistical Inference*. Springer-Verlag, New York. 207 pp.
- Walters, C., and Ludwig, D. 1994. Calculation of Bayes posterior probability distributions for key population parameters. *Canadian Journal of Fisheries and Aquatic Sciences* 51: 713-722.

Appendix 1

In this section, we provide the BUGS code used to estimate parameters of the true-catch model specified in the methods section (for scenario 1). To map this code onto the equations of that section, it is important to know that 'isigma2' corresponds to $1/\sigma^2$, 'itau2' to $1/\tau^2$, 'P[t]' to ' $B'(t)$ ', 'itheta2' to $1/\theta^2$, 'q' to q_s and 'qcomm' to q_c . Note that lines starting with '#' are comments. BUGS also uses an unconventional parameterisation for its normal and lognormal distributions: the second parameter is the reciprocal of the variance, also known as the precision of the distribution.

The ' I ' symbols following distribution declarations are used to signal the presence of censored data in BUGS. For example, ' $Y \sim dnorm(0,1)I(2,)$ ' would signify that Y has a standard normal distribution, and that there has been a censored observation ' $Y > 2$ '. Just by way of clarification, it would not mean that Y has a truncated normal distribution, which is subtly different. Changing that suffix to read ' $I(,3)$ ' would imply an observation of ' $Y < 3$ ', while ' $I(2, 3)$ ' would mean an observation of ' $2 < Y < 3$ '. These I symbols are the key to implementing censored data models in BUGS.

```

model;
{
  #priors
  r~dlnorm(-1.4,4)
  isigma2 ~ dgamma(100,.05)
  K ~ dlnorm(6.551,4)

  # _ A _
  # itheta2 ~ dgamma(70,1)
  # _ A' _

  itau2 ~ dgamma(44,2)
  qcomm~dbeta(5.5,44.5)
  q~dbeta(4,8)
  Pinit ~dbeta(20,6)

  #Dynamics
  Pmed[1] <-Pinit

  for( t in 1 : N ) {
    P[t] ~ dnorm(Pmed[t],isigma2)I(0.001,)

    # _ B _
    C[t] <-Landings[t]
    # _ B' _

    Pmed[ t + 1 ] <- ( r + 1 ) * P[t] - ( r * P[t] ) * P[t] - C[t] / K

    lmed[t] <- log((K * P[t]) * q)
    l[t] ~ dlnorm(lmed[t],itau2)
  }

  Pcur ~dnorm(Pmed[N+1],isigma2)I(0.001,)

  X<-Pcur/.5
  MSY<- r * K /4
}

```

BUGS needs data and inputs to start an MCMC chain. A typical data file would look as follows:

```
list(N=10, Landings=c(5.94,6.36,7.65,7.47,8.86,11.95,8.17,5.52,6.35,3.96),
l=c(51.17,61.25,55.67,60.07,58.31,70.99,53.8,47.54,39.73,49.09))
```

The values for ‘Landings’ were taken from a simulated data set, as were the values in the ‘l’ vector, which represents the survey index values. N is the number of years of data.

A typical input file might look as follows:

```
list(q=0.332, qcomm=0.13, itau2=25.2, isigma2=2508, Pcur=0.5, Pinit=0.537, K=193, r=0.265,
P=c(0.54,0.55,0.55,0.54,0.53,0.53,0.48,0.47,0.45,0.5))
```

An input file can be created by multiplying the true parameters by either 0.7 or 1.3. In this paper, we made two input files for every set of simulation data by using one of these two multipliers on every parameter, while making sure to constrain the parameter values in the

feasible range.

The true-catch model can be changed into the censored-catch model in four steps:

1. Remove the comment symbol '#' before 'itheta2' between points '_ A _' and '_ A'_
2. Replace the code between '_ B _' and '_ B'_ with the following:

```
Cmed[t] <- log((E[t] * qcomm) * P[t] * K)
L2[t] <= 2 * Landings[t]
C[t] ~ dlnorm(Cmed[t], itheta2) | (Landings[t], L2[t])
```

3. Add the effort data (from the same data set) to the data file:

```
list(N=10, E=c(0.42,0.49,0.55,0.62,0.68,0.75,0.66,0.57,0.48,0.39),
      Landings=c(5.94,6.36,7.65,7.47,8.86,11.95,8.17,5.52,6.35,3.96),
      l=c(51.17,61.25,55.67,60.07,58.31,70.99,53.8,47.54,39.73,49.09))
```

4. Add initial values for catch vector C for the input file:

```
list(q=0.332, qcomm=0.13, itau2=25.2, isigma2=2717,
      theta2=71, Pcur=0.5, Pinit=0.54, K=193.4, r=0.265, P=c(0.54,0.55,0.55,0.54,0.53,0.53,0.48,0.47,0.45,0.5),
      C=c(7.9,8.4,9.7,9.5,10.9,13.9,10.2,7.5,8.3,6))
```

The true-catch model can also be changed into the survey-and-effort model in four steps:

1. Remove the comment symbol '#' before 'itheta2' between points '_ A _' and '_ A'_
2. Replace the code between '_ B _' and '_ B'_ with the following:

```
Cmed[t] <- log((E[t] * qcomm) * P[t] * K)
C[t] ~ dlnorm(Cmed[t], itheta2)
```

3. Add the effort data (from the same data set) to the data file and remove the landings:

```
list(N=10, E=c(0.42,0.49,0.55,0.62,0.68,0.75,0.66,0.57,0.48,0.39),
      l=c(51.17,61.25,55.67,60.07,58.31,70.99,53.8,47.54,39.73,49.09))
```

5. Add initial values for catch vector C for the input file

```
list(q=0.332, qcomm=0.13, itau2=25.2, isigma2=2717,
      theta2=71, Pcur=0.5, Pinit=0.54, K=193.4, r=0.265, P=c(0.54,0.55,0.55,0.54,0.53,0.53,0.48,0.47,0.45,0.5),
      C=c(7.9,8.4,9.7,9.5,10.9,13.9,10.2,7.5,8.3,6))
```