A proposed model transformation to circumvent problems encountered with the sex-ratio approach to obtain lower confidence limits for the abundance of West Greenland minke whales

A. Brandão and D.S. Butterworth

MARAM (Marine Resources Assessment and Management Group)
Department of Mathematics and Applied Mathematics
University of Cape Town, Rondebosch 7701, South Africa

ABSTRACT

An illustrative example is given of a transformation applied to the sex-ratio approach which operates in a way that for population sizes much greater than are realistic, the impact of catches of females on abundance is damped. This leads to finite estimates of carrying capacity *K* even in circumstances where the trend over time in the proportion of whales in the catch that are male is decreasing (as is the case for West Greenland minke whales). The example is shown to produce positively biased estimates of the lower 5% confidence interval for current population size. However the concern with which this should be viewed is difficult to assess since the estimator is positively biased even in circumstances where the proportion of the catches that are male does not trend downwards over time.

INTRODUCTION

As reported in IWC (2010), the sex-ratio approach with its associated proposed method to obtain lower confidence limits for the abundance of the West Greenland minke whales has met with some implementation problems. This paper reports on the results for an illustrative example of one of the remedies put forward at that meeting, namely to re-parameterise the model by some suitable transformation. A simple production model is used to generate catch data that displays similar characteristics to those found in the West Greenland minke whale catch data (i.e. the proportion of the catches that are male trends downwards over time despite the fact that more females than males are being caught) and a simulation approach is used to assess the appropriateness of the proposed transformation.

METHOD

A sex-structured age-aggregated Schaefer model (similar to that described in Brandão and Butterworth (2008)) is used to demonstrate the behaviour of the proposed transformation. The description of the simulation algorithm and the catch data generation process is given in the Appendix.

Population dynamics

A sex-structured age-aggregated (or production) model is used:

$$N^m_{y+1} = N^m_y + \frac{r}{2} N^f_y \left(1 - \left(\frac{N_y}{K}\right)\right) - C^m_y \quad (1)$$

$$N^f_{y+1} = N^f_y + \frac{r}{2} N^f_y \left(1 - \left(\frac{N_y}{K}\right)\right) - \gamma(N^f_y) C^f_y \quad (2)$$

where

N_y is the total number of minke whales at the start of year y, which is given by:

$N_y = N^m_y + N^f_y$,

N^m_y is the total number of male minke whales at the start of year y,

N^f_y is the total number of female minke whales at the start of year y,

K is the carrying capacity,

C^m_y is the number of male whales caught in year y,

C^f_y is the number of female whales caught in year y,

r is the intrinsic population growth rate, which is linked to the assumption of a 50:50 sex ratio at birth; in this application r is set to 0.04, and

$\gamma\left(N^f_y\right)$ is a function of N^f_y that tends towards zero for values of N^f_y well above realistic values of $K/2$, and is given by:

$$\gamma\left(N_y^f\right)=\frac{1}{1+e^{\left(N_y^f-N^*\right)/\delta}},$$

where N^* is set at a value much larger than K could be in reality (N^* is set to 100 000 in this instance) and δ is set to be equal to $0.1N^*$.

The number of male and female whales is assumed to be the same before exploitation so that $N_1^m = N_1^f = K/2$.

The expected number of female whales caught is given by:

$$\hat{C}_y^f = C_y \frac{N_y^f}{N_y^f + \lambda N_y^m}, \tag{3}$$

where

λ^i is the selectivity of males relative to females, and is assumed to remain constant, with equation (3) following from the associated assumptions that:

$$\hat{C}_y^f = F_y N_y^f;\ \hat{C}_y^m = \lambda F_y N_y^m. \tag{4}$$

The transformation proposed (the introduction of the $\gamma\left(N_y^f\right)$ factor in equation (2)) generalises the population dynamic equations (equations (1) and (2)) so that they manifest the following features:

a) The equations remain effectively unchanged from their standard form in the range of population sizes that are realistic.
b) They admit the possibility that even if the number of females caught is greater than that of males, so that the ratio of male to female catches would be expected to increase over time, this could also decrease. Thus a decreasing trend in this ratio will still yield a finite MLE for K, hence accommodating the feature of the actual data for West Greenland minke whales that is otherwise the source of the original problem.
c) They keep the number of males and the number of females at the start of harvesting finite and positive.

d) They are continuous and differentiable across the full feasible parameter space, thus admitting a likelihood profile basis to obtain confidence limits as well as allowing the use of ADMB.

The likelihood function

The likelihood is calculated assuming that the observed female catches are distributed about their expected value according to an overdispersed Poisson model. The negative of the approximate log-likelihood (ignoring constants) which is minimised in the fitting procedure is thus given approximately by:

$$-\ln L = \sum_y \left\{ \frac{1}{2\sigma^2} \frac{\left(C_y^f - \hat{C}_y^f\right)^2}{\hat{C}_y^f} + \ln\sigma + \ln\sqrt{\hat{C}_y^f} \right\} \qquad (5)$$

where

σ measures overdispersion of the distribution of catches compared to a Poisson distribution for which the variance is equal to the expected catch, whose maximum likelihood estimate is given by:

$$\hat{\sigma} = \sqrt{\frac{1}{n} \sum_y \left\{ \frac{\left(C_y^f - \hat{C}_y^f\right)^2}{\hat{C}_y^f} \right\}} \qquad (6)$$

n is the total number of years in the summation.

Note that the formulation of equation (5) assumes that the Poisson-like catch distribution can be approximated by a normal distribution of the same variance. The estimable parameters of this model are λ, σ and K.

RESULTS AND DISCUSSION

Two sets of generated catch data are considered in this paper. One ensures that there are some simulations that will contain catch data that will have the proportion of catches that are male decreasing over time (i.e. have a negative slope) even though female catches increase (13 out of 100 in this application – the value of n^* (see Appendix) was set to 20). The second set, used as sensitivity test, is constructed so as not to contain any generated catch data with a negative slope ($n^* = 140$).

Table 1 gives the true values for the carrying capacity (*K*) and the number of minke whales at the start of the year following the period of catches considered (N_{11}). The true lower 5% confidence limits for *K* and N_{11} obtained from the distributions of estimates of these parameters are also given. Results are shown when data is generated in such a way as to allow for the possibility of negative slopes in the proportion of whales caught which are male, as well as when no such negative slopes occur. Figures 1 and 2 show the distribution of the *K* and N_{11} estimates respectively for the simulation exercise that allows for negative slopes. The histograms of the parameter estimates are also disaggregated in terms of generated data sets where these slopes are positive and where they are negative. Figure 3 shows the histogram of the estimates of *K* and N_{11} for the sensitivity test when no negative slopes occur in the catch data generation exercise.

Mean estimates, and bias and precision of the parameter values and of the lower confidence bounds are reported in Table 2 (for *K*) and in Table 3 (for N_{11}). Table 4 reports these results when no negative slopes in the proportion of whales caught which are male occur in the simulation exercise.

Figures 4 and 5 show the estimated lower 5% confidence bounds for *K* and N_{11} respectively (also split into cases generated with only positive and only negative slopes). The true 5% confidence limits are shown as an arrow in these plots. Figure 6 shows the estimated lower 5% confidence bounds for the sensitivity test with no negative slopes generated.

As would be expected, for cases where realisations of catch data showing negative slopes over time in the proportion of whales that are male can occur, both *K* and N_{11} estimates show substantial positive bias because of the large (though finite) MLEs which eventuate in such circumstances. Note that even for the sensitivity test without such negative slopes and hence no instances of very large estimates, the (effectively standard) estimator shows some positive bias (roughly 25%).

The real interest in this approach is, however, in its ability to estimate lower 5% confidence intervals, particularly for current population size (N_{11} in this example). Again the estimator is positively biased, and it is larger for the standard analysis (with negative slopes: 1513) than for the sensitivity test (without such slopes: 676). Results for *K* are similar.

IN CONCLUSION

This exploratory exercise seems promising, but is in no way definite. A difficulty in interpreting the results is that the estimator, including the likelihood profile approach to obtain the lower confidence interval, is biased even for well-behaved data (the sensitivity test). Thus it is difficult to separate the effects of this aspect of bias from those that may be associated with the possibility of negative slopes.

A fuller investigation would require consideration of a greater number of simulations and an age-structured population model (at least). This would not be trivial, as computation of the likelihood profiles in this exercise proved to be difficult (in relation to ensuring minimisation convergence), necessitating considerable extra time spent by the analyst.

ACKNOWLEDGMENT

This work was supported by the South African National Research Foundation.

REFERENCE

Brandão, A and Butterworth, D.S. 2008. Production model based estimates of lower confidence limits for the abundance of West Greenland minke whales. IWC document SC/60/AWMP9.

International Whaling Commission. 2010. Report of the Standing Working Group (SWG) on the Aboriginal Whaling Management Procedure (AWMP). Annex E.

Table 1. True values for K, N_{11} and the true lower 5% confidence limits for K and N_{11}.

		K	N_{11}
True value		10 000	6 224
Lower 5% confidence limit	**Some negative slopes in simulation exercise**	6 097	2 233
	No negative slopes in simulation exercise	7 505	3 687

Table 2. Simulation mean, standard deviation, bias and RMSE for K and the lower 5% confidence limit for K, when some of the generated catch data series show negative slopes in the proportion of the catch that is male.

	***K* estimate**			**Lower 5% confidence limit for *K***		
	All data	**Positive slope**	**Negative slope**	**All data**	**Positive slope**	**Negative slope**
Mean	44 586	12 119	261 867	7 554	7 060	10 856
Standard deviation	84 890	9 279	6 857	1 932	1 039	3 077
Bias	34 586	2 119	251 867	1 457	964	4 760
RMSE	91 271	9 466	251 953	2 412	1 413	5 603

Table 3. Simulation mean, standard deviation, bias and RMSE for N_{11} and the lower 5% confidence limit for N_{11}, when some of the generated catch data series show negative slopes in the proportion of the catch that is male.

	N_{11} estimate			**Lower 5% confidence limit for N_{11}**		
	All data	**Positive slope**	**Negative slope**	**All data**	**Positive slope**	**Negative slope**
Mean	41 145	8 333	260 731	3 745	3 229	7 201
Standard deviation	85 785	9 318	6 972	1 982	1 063	3 070
Bias	34 921	2 109	254 507	1 513	996	4 968
RMSE	92 222	9 502	254 595	2 485	1 452	5 778

Table 4. Simulation mean, standard deviation, bias and RMSE for K and N_{11} and the lower 5% confidence limit for K and N_{11}, when none of the generated catch series show negative slopes in the proportion of the catch that is male.

	Estimate		**Lower 5% confidence limit for N_{11}**	
	K	N_{11}	K	N_{11}
Mean	12 331	8 558	8 170	4 364
Standard deviation	15 035	15 075	1 293	1 316
Bias	2 331	2 335	665	676
RMSE	15 140	15 180	1 448	1 474

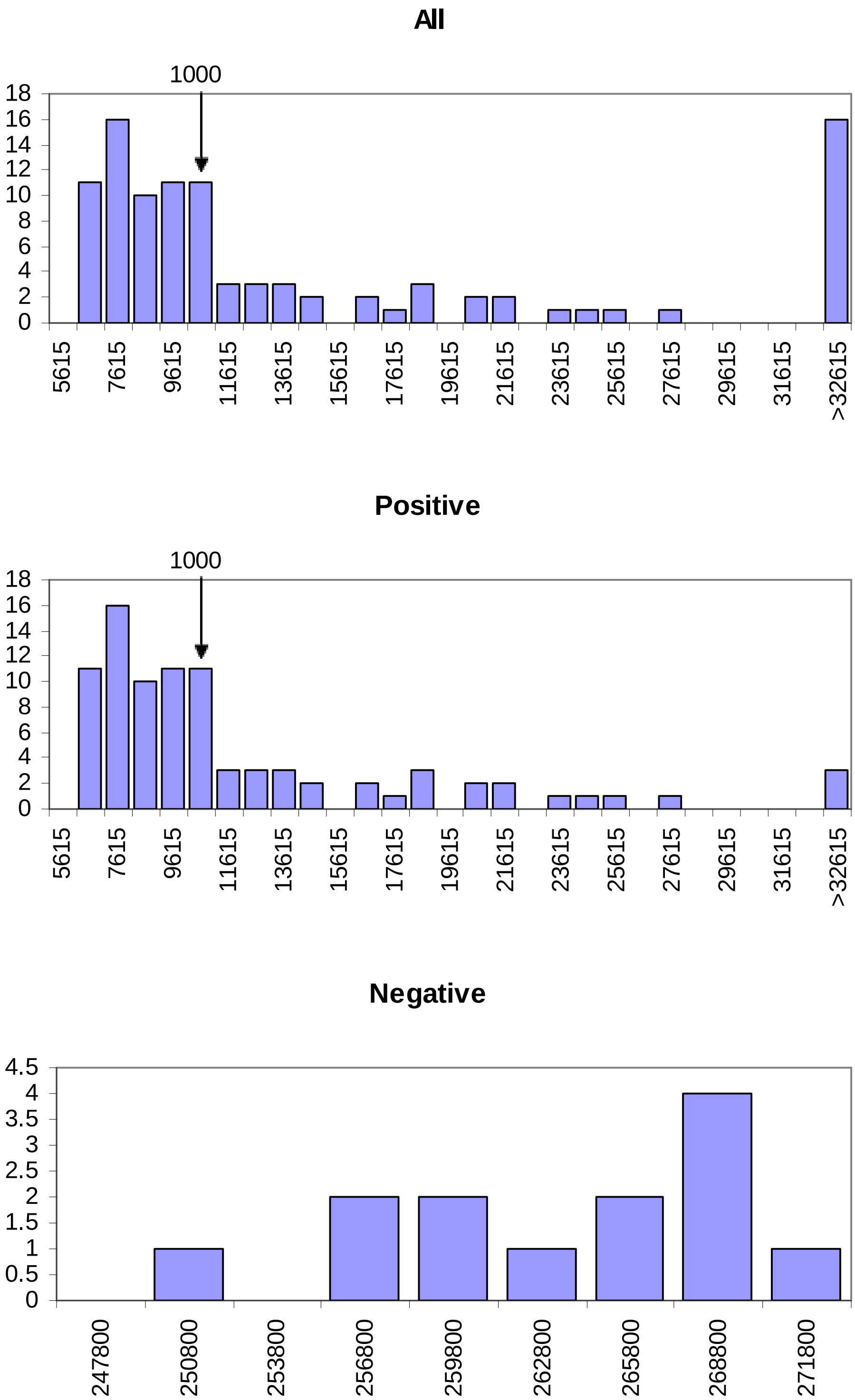


Figure 1. Histogram of *K* estimates from all simulations (top), from the catch data generated with a positive (middle) and with a negative (bottom) slope over time in the proportion of the catch made that is male. The arrow shows the true value.

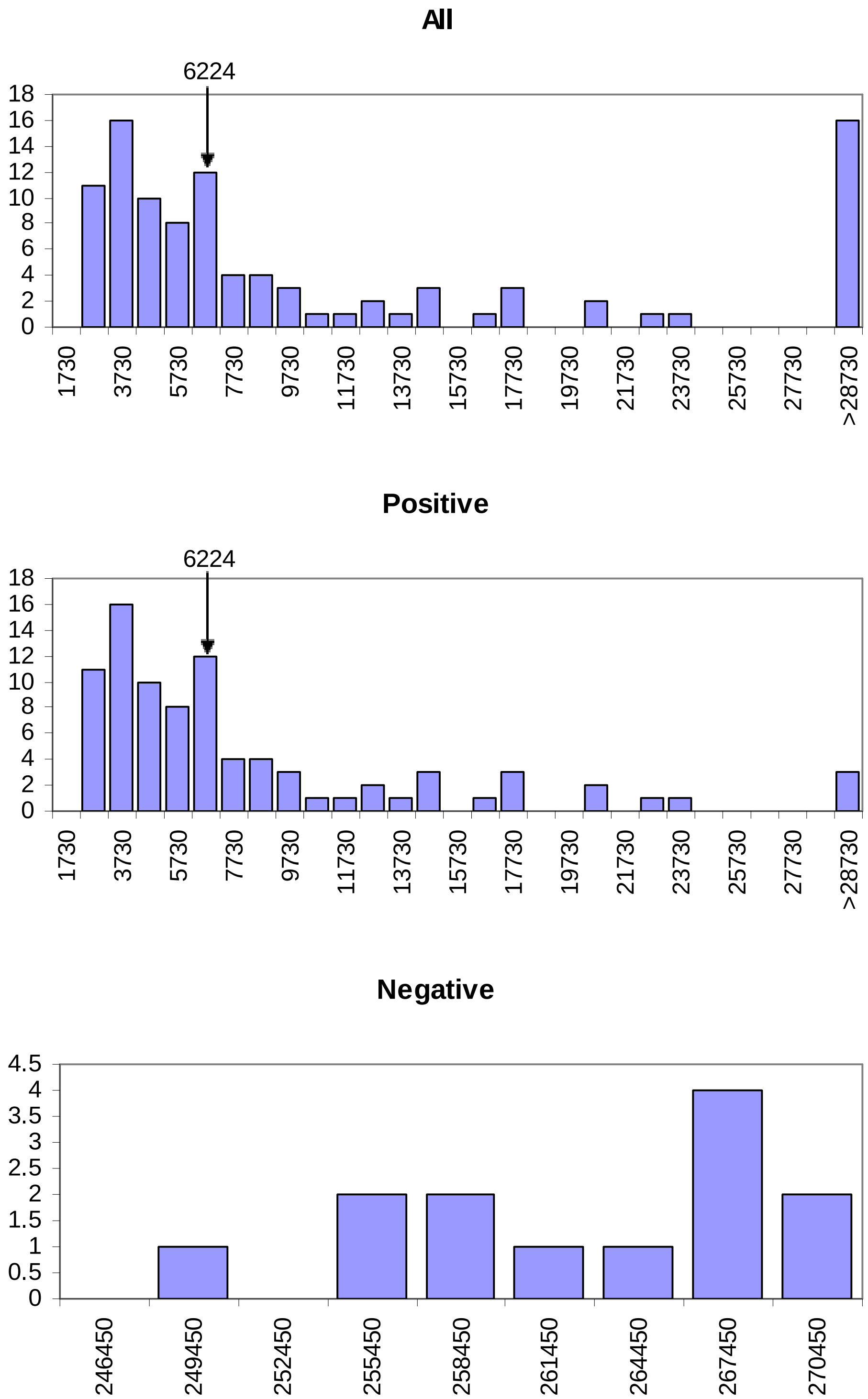


Figure 2. Histogram of N_{11} estimates from all simulations (top), from generated catch data with a positive (middle) and with a negative (bottom) slope over time in the proportion of the catch made that is male. The arrow shows the true value.

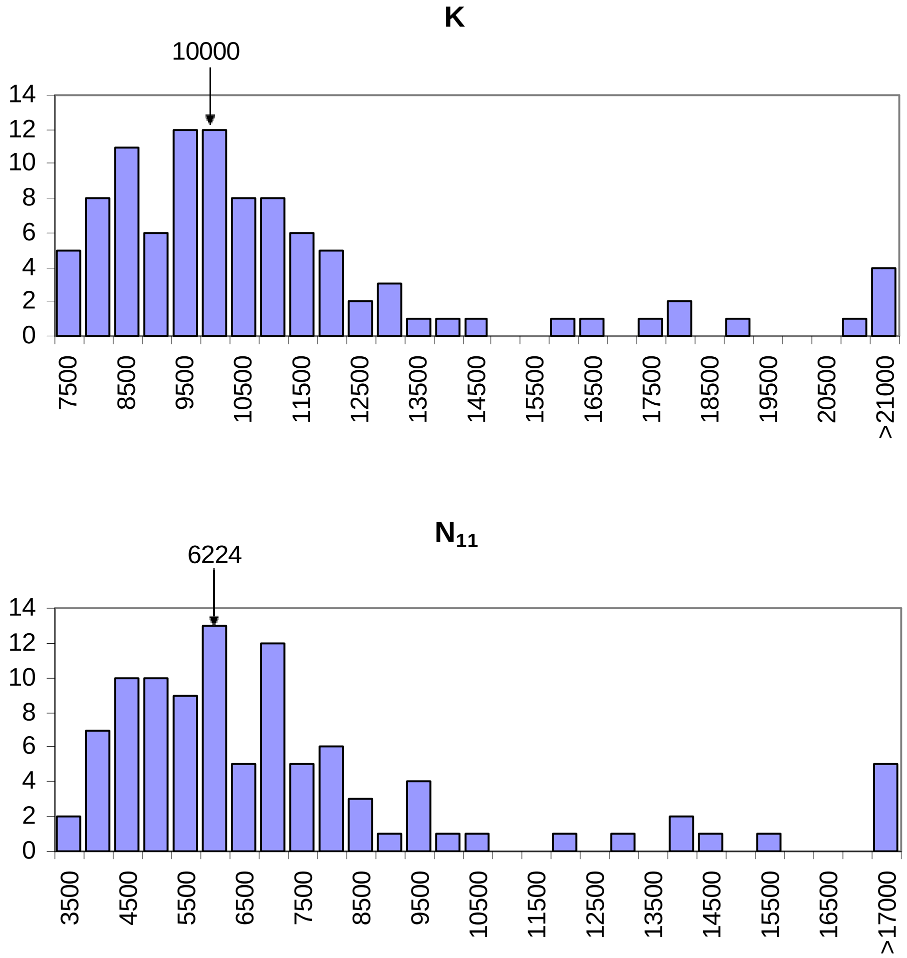


Figure 3. Histograms of K and N_{11} estimates from simulations when none of the generated catch data contain negative slopes over time in the proportion of the catch made that is male. The arrow shows the true value.

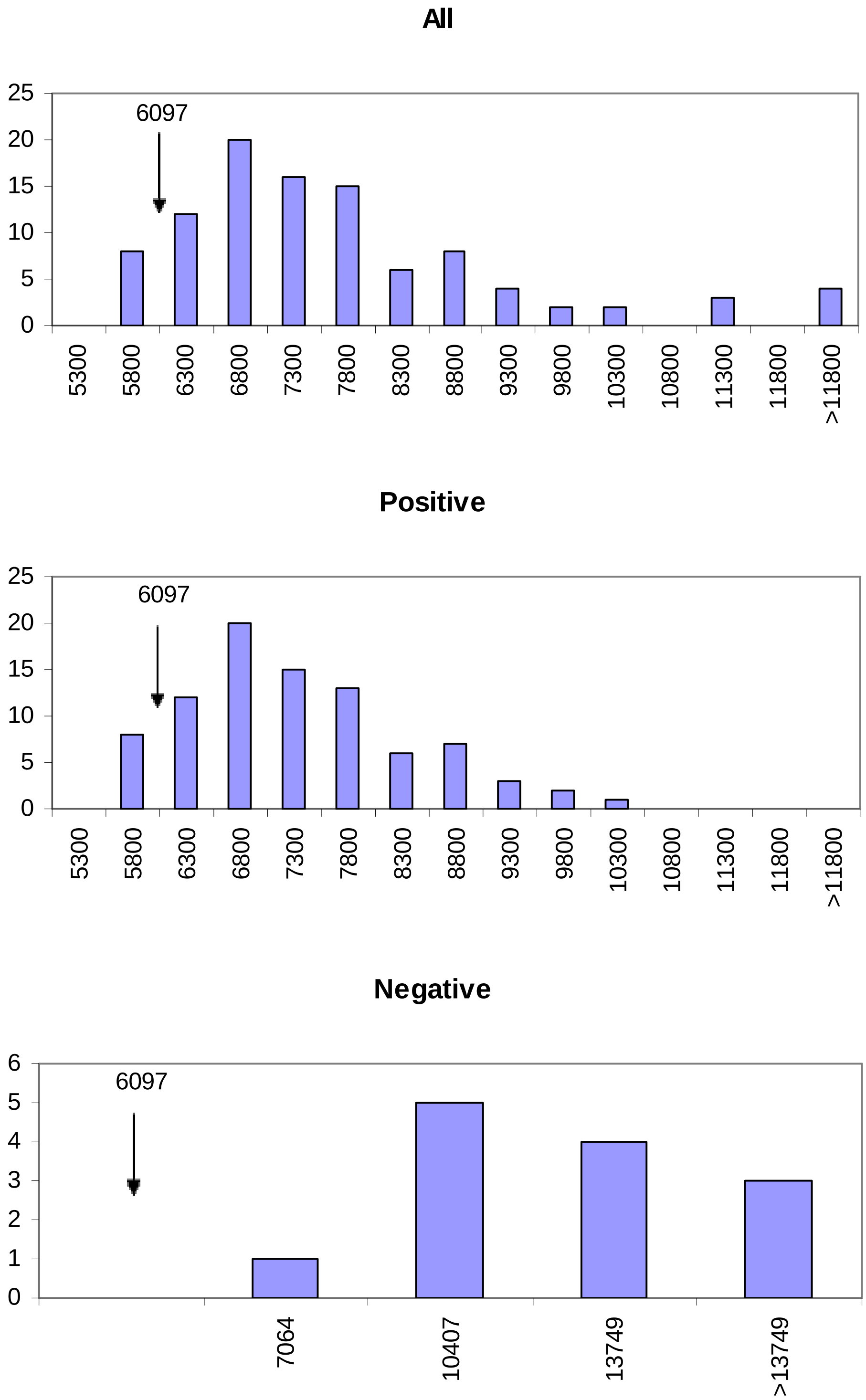


Figure 4. Histogram of lower 5% confidence limit for *K* from all simulations (top), from generated catch data with a positive (middle) and with a negative (bottom) slope over time in the proportion of the catch made that is male. The arrow shows the true 5% confidence limit.

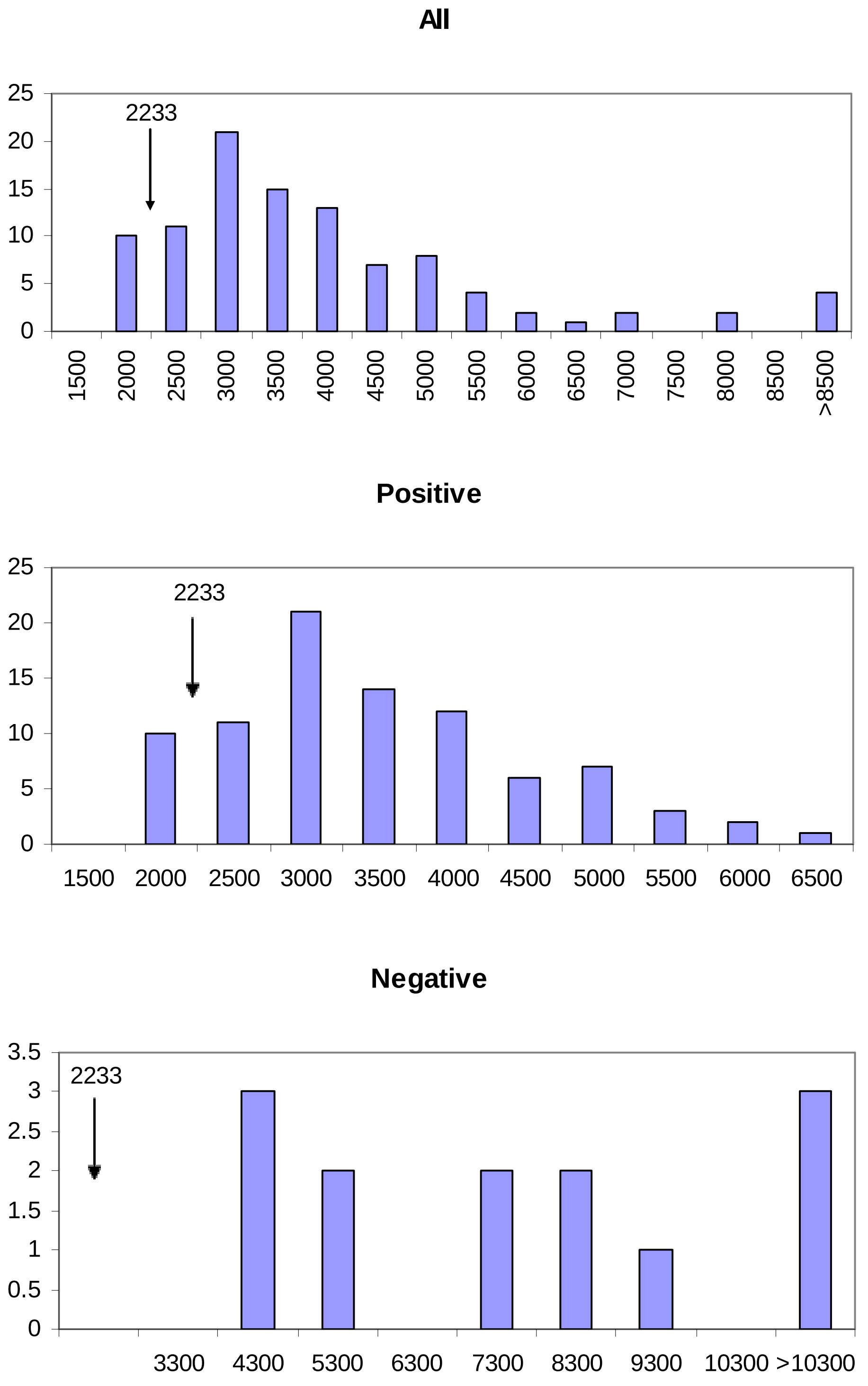
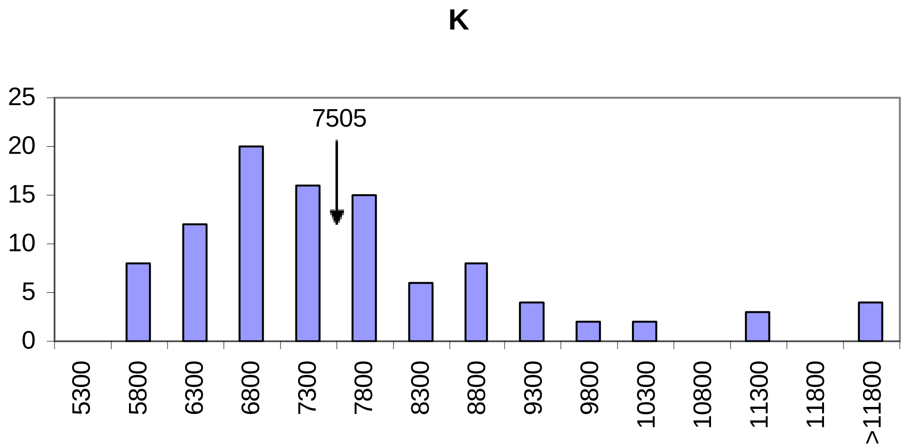


Figure 5. Histogram of lower 5% confidence limit for N_{11} from all simulations (top), from generated catch data with a positive (middle) and with a negative (bottom) slope over time in the proportion of the catch made that is male. The arrow shows the true 5% confidence limit .



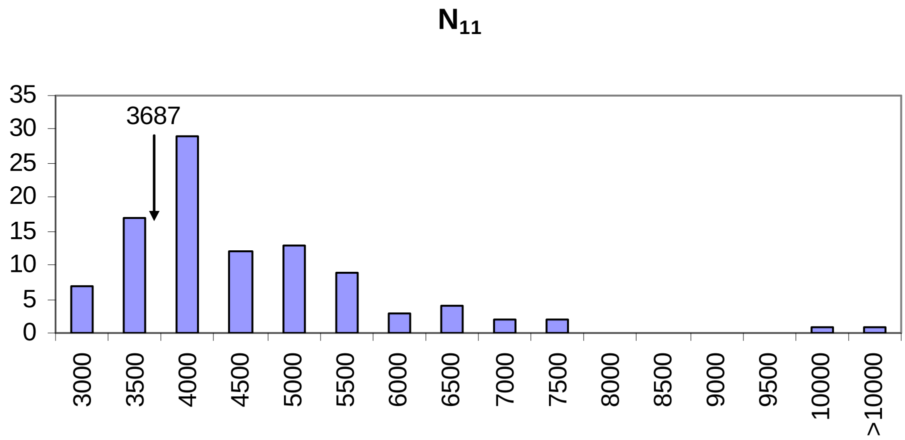


Figure 6. Histogram of lower 5% confidence limit for *K* and N_{11} from simulations when none of the generated catch data contain negative slopes over time in the proportion of the catch made that is male. The arrow shows the true 5% confidence limit.

APPENDIX

Simulation algorithm

The true value of virgin biomass (K) is fixed at 10 000. For this set value of K and the same total annual catches each year ($C_y = C_y^m + C_y^f = 400$), the following steps are taken:

1. Set $N_1 = K$; $N_1^m = N_1^f = K/2$.
2. Generate C_y^f, C_y^m (as described below).
3. Given C_y^f and C_y^m, project N_y^f and N_y^m forward one year (using equations (1) and (2)).
4. Repeat steps (2) and (3) until the end of the time period (i.e. here $y = 10$).
5. Fit the model to the data generated by minimising the negative log-likelihood function of equation (5) to obtain estimates of K and N_{11} for these generated data.
6. Obtain the lower 5% profile likelihood value for both the K and N_{11} estimates.
7. Repeat steps (1) to (6) 100 times to get the distribution of the estimates of K and N_{11} as well as the distribution of the lower 5% profile likelihood values for K and N_{11}.

Data generation

Data are generated assuming that K = 10 000 and that the total catch each year remains constant (and assumed to total 400 in this application). Data are generated for a period of 10 years. The expected number of females caught is assumed to be given by:

$$\hat{C}_y^f = C_y \frac{N_y^f}{N_y^f + 1/3 N_y^m} \qquad \text{(A.1)}$$

so that the probability of a male being caught is given by:

$$\hat{p}_y^m = \frac{N_y^m}{N_y^m + 3N_y^f} . \qquad \text{(A.2)}$$

However, to include overdispersion in the data generated, the number of males caught in a particular realisation, $\hat{C}_y^{m(r)}$, is obtained by drawing $p_y^{m(r)}$ from a binomial distribution with parameters n^* and $\hat{p}_y^m$, where n^* is chosen to be much less than 400 and in such a way that within the simulated data sets, some of the time series for the proportion of the catch that is male will decrease (in this application 13 out of 100 simulations showed negative slopes against year in the proportion of the catch that is male). A sensitivity test was carried out in which n^* was chosen so that no simulations had a negative slope. The number of males caught in the data sets generated is given by:

$$C_y^m = C_y p_y^{m(r)}, \tag{A.3}$$

and thus corresponding number of females caught is given by:

$$C_y^f = C_y - \hat{C}_y^m. \tag{A.4}$$

The true value of N_{11} is obtained by projecting N_y^f and N_y^m forward using equations (1) to (2) where the number of males caught is given by equation (A.3) and the number of females by equation (A.4).

Lower 5% confidence limit

The true lower 5% confidence limits for K and N_{11} are obtained as the 5th percentiles of the distributions of the estimates of K and N_{11} respectively. The distributions of the estimated lower 5% confidence limits for K and N_{11} are obtained from the profile likelihood estimates for each simulated data set.

Bias and precision of estimators

The mean (and the standard deviation) of the estimates for K and N_{11} are obtained from the distributions of the K and N_{11} estimates. That is, for the estimator of K, the mean estimate is given by $\bar{K} = \frac{1}{S}\sum_s \hat{K}_s$ and the standard deviation of the estimate is given by $SD = \sqrt{\frac{1}{S-1}\sum_s (\hat{K}_s - \bar{K})^2}$ respectively, where S is the number of simulations performed. The bias of the estimator for K is then given by $\bar{K} - K$, where K is the true value for the parameter K

(fixed at 10 000 in this application). The RMSE of the *K* estimator is given by $RMSE = \sqrt{\frac{1}{S} \sum_s (\hat{K}_s - K)^2}$.