Further results for 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

The methodology advanced by Brandão and Butterworth (2007 and 2008) for simulating catch sex sampling data in the process of computing lower confidence bounds for the abundance of the West Greenland minke whale population is applied to the new modelling scenarios developed at the March 2008 IWC Workshop on Greenlandic Fisheries. To date calculations have extended only to MSYR = 2%, and been implemented for Models 3, 4b and 5 for the age-structured population model of Witting and Schweder (2008). For Model 5 for the SW stratum treated as an independent stock, the MLE of pre-exploitation abundance K is finite, so that standard likelihood profile estimates of lower confidence bounds for abundance are provided. Lower bounds for Model 4b (the "influx" model) are higher than for Model 3 (the "redistribution" model. Non-convergence of some of the Model 3 fits required to compute the deviance distributions which are used to determine the confidence bounds still persist, but time constraints have precluded further investigation of suggestions put forward in IWC (2008b).

INTRODUCTION

This paper reports applications of the methodology set out in IWC (2008a) for the computation of lower confidence limits for the abundance of West Greenland minke whales using an assessment model based upon imbalanced catch sex-ratio data and under the procedures of Brandão and Butterworth (2008) for simulating such catch sex sampling data. The scenarios postulated in Models 3, 4b and 5 of IWC (2008) are investigated using an age structured model (identical to that in Witting and Schweder, 2008) for the resource dynamics. This paper addresses some of the suggestions set out in IWC (2008b) to deal with problems of model convergence and negative deviances encountered in Brandão and Butterworth (2008).

DATA

The data used for these analyses are generally identical to those used by Witting and Schweder (2008) but with additional 2007 catches as provided by Witting (*pers. comm.*). For Model 5 the prescriptions of IWC (2008a) were used; those were incomplete in one respect (the specification for the split of the unsampled whales by stratum and sex for the early Greenlandic period of whaling; the procedures used to specify these splits are set out in the Appendix).

METHODS

The basic methodology used is the same as that of Brandão and Butterworth (2008) but in this paper only an age-structured population model is used. The population dynamics model is identical to that described in Witting and Schweder (2008). The description of the simulation algorithm and the data generation process is given in the Appendix. Evaluations to deal with new scenarios specified by IWC (2008) are detailed further below; the basic approach to compute deviance distributions has followed that set out in IWC (2008a), and is not repeated in detail here.

Population dynamics
A sex-structured age-structured model is used:

$$N_{y+1,0}^g = 0.5 \rho_{y+1} N_{y+1}^{m(f)} \tag{1}$$

$$N_{y+1,a+1}^g = (N_{y,a}^g - C_{y,a}^g) s_a \qquad 0 \le a \le m\text{-}2 \tag{2}$$

$$N_{y+1,m}^g = (N_{y,m}^g - C_{y,m}^g) s_a + (N_{y,m-1}^g - C_{y,m-1}^g) s_{a-1} \tag{3}$$

where:

g is the gender (male/female) of the minke whale,

$N_{y,a}$ is the total number of minke whales of age a in year y, which is given by:

$$N_{y,a} = N_{y,a}^m + N_{y,a}^f,$$

$N_{y,a}^{g}$ is the total number minke whales of gender g of age a in year y,

$C_{y,a}^{g}$ is the number of West Greenland minke whales of gender g of age a caught in year y,

ρ_{y+1} is the fecundity rate for year $y+1$, which takes the Pella_Tomlinson form:

$$\rho_{y+1} = b_k + \left[b_{\max} - b_k \right] \left[1 - \left(N_{y+1}^{1+} / K^{1+} \right)^{z} \right], \text{ where}$$

         b_k is the birth rate at carrying capacity K,

         b_{max} is the maximal birth rate,

         z is the strength of density dependence (the degree of compensation parameter).

         N_{y+1}^{1+} is the total number of the one plus component of the minke whales population in year $y+1$ given by:

$$N_{y+1}^{1+} = \sum_{a=1}^{m} N_{y+1,a}^{f} + N_{y+1,a}^{m},$$

         K^{1+} is the carrying capacity of the one plus component of the minke whales population given by:

$$K^{1+} = \sum_{a=1}^{m} N_{1948,a}^{f} + N_{1948,a}^{m},$$

$N_{y+1}^{m(f)}$ is the number of mature females at the start of year $y+1$, given by:

$$N_{y+1}^{m(f)} = \sum_{a} \gamma_a N_{y+1,a}^{f}, \text{ where}$$

$$\delta_a = \begin{cases} 0 & a < a_m \\ 1 & a \geq a_m \end{cases}, \text{ where } a_m \text{ is the age of reproductive maturity,}$$

m is the maximum age considered (i.e. the "plus group").

s_a is the age specific annual survival rate, given by:

$$s_a = \begin{cases} s_{juv} s_{ad} & a = 0 \\ s_{juv} & 1 \leq a \leq a_{ad} \\ s_{ad} & a > a_{ad} \end{cases}, \text{ where:}$$

     s_{juv} is the survival rate for juveniles,

     s_{ad} is the survival rate for adults, and

     a_{ad} is the greatest age at which the "juvenile' survival rate applies (taken as $a_{ad} = 1$).

The number of whales of gender g of age a caught in year y is given by:

$$C_{y,a}^{g} = C_{y}^{g} \frac{G_a N_{y,a}^{g}}{\sum_{a'=0}^{m} G_a N_{y,a'}^{g}} \tag{4}$$

where:

     G_a is the age-specific differentiation of the catch relative to the age composition of the overall population, given by:

$$G_a = \begin{cases} 0 & a = 0 \\ a/a_c & a < a_c \\ 1 & a \geq a_m \end{cases}, \text{ where } a_c \text{ is the age at first capture, and}$$

2

C_y^g is the total catch of minke whales of gender g in year y.

The initial numbers for each gender g at each age a are taken to follow an unexploited equilibrium distribution evaluated as follows:

$$N_{1948,a}^g = \frac{1}{2} \eta N_a^* \tag{5}$$

where:

$$\eta = \frac{K}{\sum\limits_{a=0}^{m} N_a^*},$$

$$N_a^* = \begin{cases} 1 & a = 0 \\ s_{a-1} N_{a-1}^* & 0 < a \leq m-1. \\ \dfrac{s_m N_{m-1}^*}{1 - s_m} & a = m \end{cases}$$

The total number of minke whales caught in the whaling period i in year y is given by:

$$C_y^i = C_y^{m(i)} + C_y^{f(i)} \tag{6}$$

where:

$C_y^{m(i)}$ is the number of male minke whales caught in period i by the fishery concerned, where the period/fishery i represents:

$$i = \begin{cases} I & \text{the period } 1955 - 1978 \text{ by Greenlanic whalers} \\ II & \text{the period } 1968 - 1985 \text{ by Norwegian whalers , and} \\ III & \text{the period } 1987 - 2006 \text{ by Greenlanic whalers} \end{cases}$$

$C_y^{f(i)}$ is the corresponding number of female minke whales caught in period/fishery i.

For the simplest form of the model (Model 5), the expected number of female minke whales caught by each period/fishery i in year y is given by:

$$\hat{C}_y^{f(i)} = C_y^i \frac{N_y^f}{N_y^f + \lambda^i N_y^m}, \tag{7}$$

where:

λ^i is the selectivity of males relative to females for the period and fishery concerned, and is assumed to remain constant over that period, with equation (6) following from the associated assumptions that:

$$\hat{C}_y^{f(i)} = F_y^{(i)} N_y^f ; \quad \hat{C}_y^{m(i)} = \lambda^i F_y^{(i)} N_y^m . \tag{8}$$

3

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 by:

$$-\ln L = \sum_{i=I}^{III} \sum_{y=y_i^1}^{y_i^*} \left\{ \frac{1}{\sigma_i^2} \frac{\left(C_y^{f(i)} - \hat{C}_y^{f(i)} \right)^2}{\left(2\hat{C}_y^{f(i)} \right)} + \ln \sigma_i + \ln \sqrt{\hat{C}_y^{f(i)}} \right\} \tag{6}$$

where

y_i^1 is the first year of catches for period i,

y_i^* is the last year of catches for period i,

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

$$\hat{\sigma}_i = \sqrt{\frac{1}{n_i} \sum_{i=I}^{III} \sum_{y=y_i^1}^{y_i^*} \left\{ \frac{\left(C_y^{f(i)} - \hat{C}_y^{f(i)} \right)^2}{\hat{C}_y^{f(i)}} \right\}} \tag{7}$$

n_i is the total number of years in the summation of each whaling period.

Note that the formulation of equation (6) assumes that the Poisson-like catch distribution can be approximated by a normal distribution of the same variance. The estimable parameters of this basic model are λ^I, λ^{II}, λ^{III}, σ_I, σ_{II}, σ_{III} and K.

Recent refinements

The approach above is applied to three scenarios set out in IWC (2008):

Model 3: A closed population model with time dependence of the fractions of females and of males distributed in the NW + CW and SW strata in the later period of Greenlandic whaling commencing in 1987. To reflect this variation, the proportions of males and of females in the NW + CW stratum (relative to the NW+CW and SW strata combined) are assumed to change with time during the recent period of Greenlandic whaling as:

$$r_t^{m/f} = e^{\alpha^{m/f} + \beta^{m/f} t} / \left[1 + e^{\alpha^{m/f} + \beta^{m/f} t} \right]$$

with a complementary change in the SW stratum. The time t is specified by the standard calendar year.

Model 4b: An "influx" model where the number of male whales in the SW stratum is assumed to be influenced by whales moving in from other areas in recent years. To effect this, the λ parameter is assumed to vary linearly from 1987 $\left(\lambda_{87} \right)$ to 2006 $\left(\lambda_{06} \right)$ when these Greenlandic catches were sampled for sex.

Model 5: The animals in the NW + CW and SW strata are assumed to exhibit site fidelity, and so (*in extremis*) are treated as separate populations with different values of λ for each period and stratum which do not change within those periods.

In other changes from earlier analyses, catch data from the Norwegian fishery period II are not taken into account in the likelihood (except for Model 5 for the NW + CW stratum), but are included amongst the catches taken into account in the population dynamics equations.

Dealing with non-convergence and negative deviances

Brandão and Butterworth (2008) reported that during the process of developing the deviance distributions, there were instances when the minimisation to obtain the best value of the likelihood had not converged and/or negative values for deviance occurred. This was a problem for Model 4b at low K values and for Model 3 at all K values, and likely relates to the likelihood being fairly flat in some of the parameters being estimated. In this paper the approach taken as an

initial attempt to overcome convergence problems has been to lower the convergence criterion of the likelihood gradient (previously convergence was set at a gradient less than 0.00001 and now this has been set to 0.001). If the gradient is greater than 0.001, the model fit is re-run with initial parameter values set to be those obtained in the last run. In cases where non-convergence still occurs, the approach has been to carry out more than 1000 so that the first 1000 converged simulations were used in calculating the quantiles.

To attempt to deal with cases where negative deviances occur, the approach has been to initialize the search over parameter values for the model that estimates K (i.e. the larger model) with the parameter estimates obtained from the model that sets K as fixed (i.e. the smaller model, which is a special case of the larger model). With this approach the parameter search for the larger model starts from position with a likelihood value equal to that of the smaller model, and thus with an additional estimable parameter K will either better the likelihood value or at worst, remain the same.

RESULTS

The core results obtained for lower confidence bounds (for N_{2007}, K and their ratio) are reported in Table 1, with the best estimates of the λ, α, β and σ parameters listed in Table 2, and the process used to determine the bounds from the deviances quartiles illustrated in Figures 1-4. Note that restrictions on time have precluded applications for other than the MSYR = 2% case.

In the case of Model 5 for the SW stratum treated as if it were an isolated stock, a finite MLE of K results, so that the simulation approach is not needed to determine lower confidence bounds, and these are instead reported using the standard likelihood profile approach.

The problem of negative deviances does not occur in any of the models with the initialisation procedure adopted here. However, non-convergence in Model 3 is still a problem. For example, for the simulation when K is 17 600 (i.e. the 5% lower confidence limit), in 1000 replicates, there were 61 cases which failed to converge for the model when K is fixed, and 43 for the model when K is estimated. If the convergence criterion was lowered to 0.01, this would drop the number of non-convergence to 28 and 24 respectively.

Comparing the different scenarios, production model results yield lower values for lower bounds on K (and for N_{2007}) for Model 3 ("redistribution") than for Model 4b ("influx"). If results for Model 5 for the two strata are added, they are intermediate between those for Models 3 and 4b.

CONCLUDING REMARKS

Due to time constraints, other suggestions set out in IWC (2008b), have not been investigated yet. The following suggestions still remain to be addressed:

High priority topics
Model 3 appears to be over-parameterised and this may be a partial explanation for the convergence problems noted in SC/60/AWMP9. An exploration of the data and alternative models should be undertaken to identify a more parsimonious parameterisation.

Construct profile likelihoods for each of the parameters of the model in turn for specific choices of K to further examine the behaviour of the estimation framework.

Examine methods to deal with the technical issues related to convergence. Results exploring these technical issues should focus on cases shown to in SC/60/AWMP1 and SC/60/AWMP9 to lead to the greatest number of convergence problems (addressed in part in this paper).

Review the approaches for handling the additional variation and examine the reasons for the differences in results for the two approaches identified at the intersessional AWMP workshop (see SC/60/Rep2 item 2.2).

Examine whether the extent of variation in the sex ratio of the catch is captured reliably and, if not, modify the approach used to generate the pseudo data sets and ensure that this is the case.

Conduct a sensitivity test that involves ignoring, in turn, the early Greenlandic data for each stratum; these sensitivity tests will allow the impact of the apparent residuals patterns to be examined.

The computer programs used to implement methods should be validated by the Secretariat (although two groups independently implementing essentially the same set of specifications should enhance confidence that the methods are correctly implemented).

Optional topics
Explore the use of a beta-binomial likelihood function.

Conduct a sensitivity test in which female fraction in model 4b relates to temperature rather than year.

In Brandão and Butterworth (2007), it was found that for some series, the procedure in the Appendix used to generate the sex-sampled catch yielded results with less variability (in median terms) than the original data, and autocorrelation was introduced into the resampling process to correct for that. This matter is being investigated further for the new scenarios; in the meantime, results reported in this document all correspond to setting that autocorrelation to zero.

ACKNOWLEDGMENTS

REFERENCES

Brandão, A and Butterworth, D.S. 2007. Further results from an AWMP-lite-like assessment of the West Greenland minke whale population. IWC document SC/59/AWMP8.

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. 2008a. Report of the Workshop on Greenlandic Fisheries.

International Whaling Commission. 2008b. Report of the Scientific Committee. *Rep. int. Whal. Commn. Annex E*.

Witting, L. and Schweder, T. 2008. Lower bound on population status from catch sex ratio: applied to minke whales off West Greenland. IWC document SC/60/AWMP1.

Table 1

Lower 5%, 10% and 50% quantiles for the population abundance in 2007 (N_{2007}), the carrying capacity (K) and the current depletion (N_{2007}/K) at the start of 2007 for various models for a MSYR value of 2%. For Model 5 (SW) results are for the likelihood profile method and hence are shown in italics (the values in the 50% columns are then the MLEs). The methods are applied to provide the quantiles for K, and then the corresponding values of N_{2007} (and N_{2007}/K) follow from the MSYR value assumed and the actual catches made. For the age-structured population model, values given are for the 0+ (total) population.

| N_{2007}				K				N_{2007}/K			
2.50%	5%	10%	50%	2.50%	5%	10%	50%	2.50%	5%	10%	50%
Model 3											
4,527	6,174	11,436	139,125	16,700	17,600	20,900	143,900	0.271	0.351	0.501	0.954
Model 4											
7,384	9,174	16,925	142,334	18,300	19,400	25,000	147,100	0.403	0.473	0.677	0.968
Model 5 (NW+CW)											
5,742	11,920	18,297	145,115	13,400	17,700	23,100	149,900	0.428	0.673	0.792	0.978
Model 5 (SW) - MLE											
463	477	494	615	4,265	4,273	4,282	4,345	0.109	0.112	0.115	0.142

Table 2.

Parameter estimates for overdispersion and male selectivity relative to females for the case of $K \rightarrow \infty$ (approximated numerically here by $K = 200\,000$) for various models for a MSYR value of 2%. Model 5 (SW) results correspond to genuine MLEs and hence are shown in italics.

Parameter	Model 3	Model 4b	Model 4b (age structured)	Model 5 (NW+CW)	Model 5 (SW)
λ^{I} (Greenland 1955-1978)	0.356	0.356	0.356	0.394	*0.243*
λ^{II} (Norwegian)	—	—	—	0.372	—
λ^{II} (Greenland NW+CW 1987-2006)	0.400	0.359	0.363	0.360	—
λ^{III} (Greenland SW 1987-2006)	0.001	—	—	—	*0.124*
λ_{87} (Greenland SW 1987-2006)	—	0.122	0.120	—	—
λ_{06} (Greenland SW 1987-2006)	—	0.389	0.379	—	—
α^{f}	178.35	—	—	—	—
α^{m}	309.80	—	—	—	—
β^{f}	-0.085	—	—	—	—
β^{m}	-0.154	—	—	—	—
σ_{I} (Greenland 1955-1978)	0.814	0.814	0.814	0.738	*0.422*
σ_{II} (Norwegian)	—	—	—	1.695	—
σ_{II} (Greenland NW+CW 1987-2006)	0.636	0.662	0.673	0.662	—
σ_{III} (Greenland SW 1987-2006)	0.364	0.370	0.378	—	*0.363*
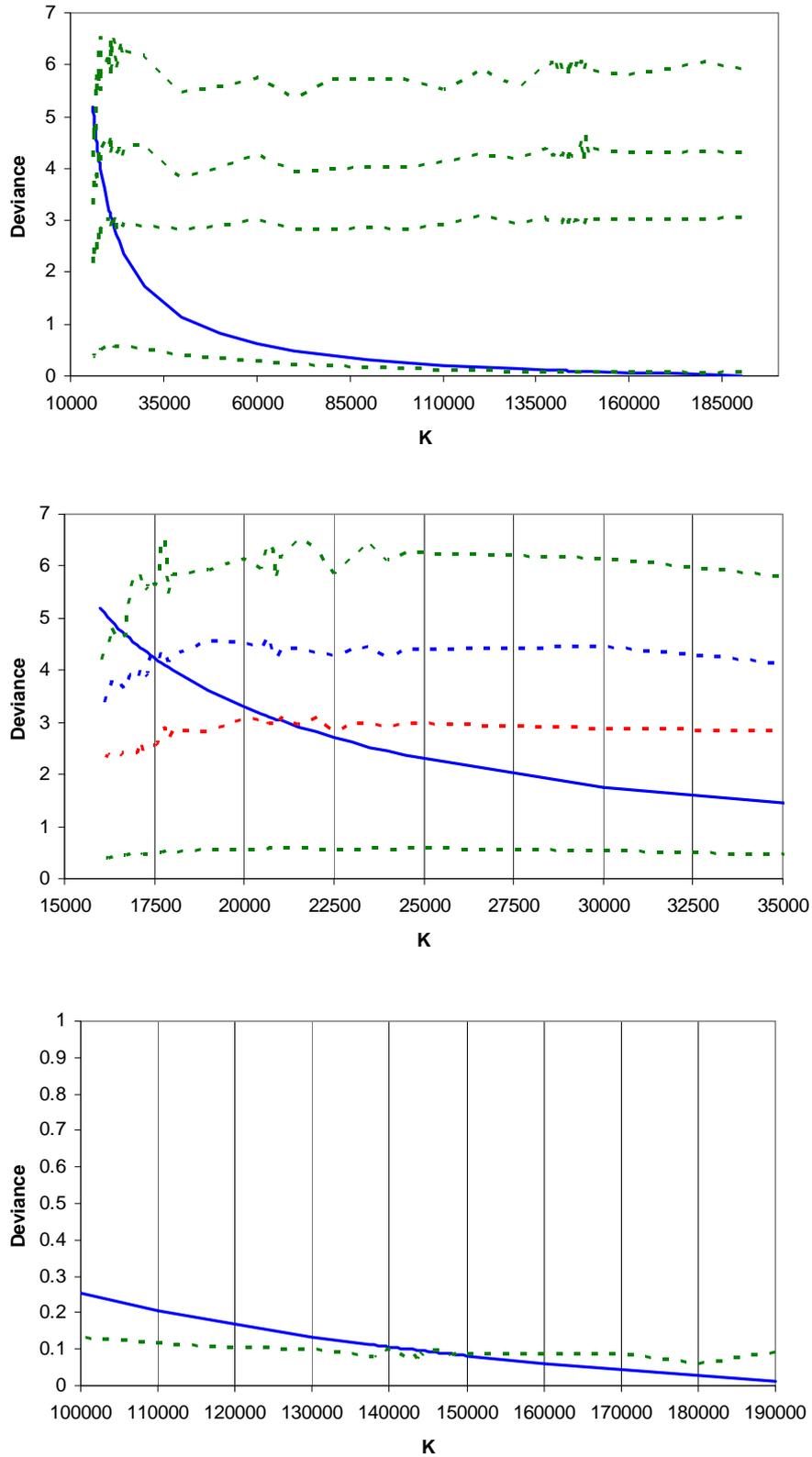
Figure 1. Observed deviance (solid line) and 2.5%, 5%, 10% and 50% quantiles (dashed lines) based on 500 simulations, together with close ups of the intersections of the quantiles and the observed deviance curve, for **Model 3** with MSYR = 2%.
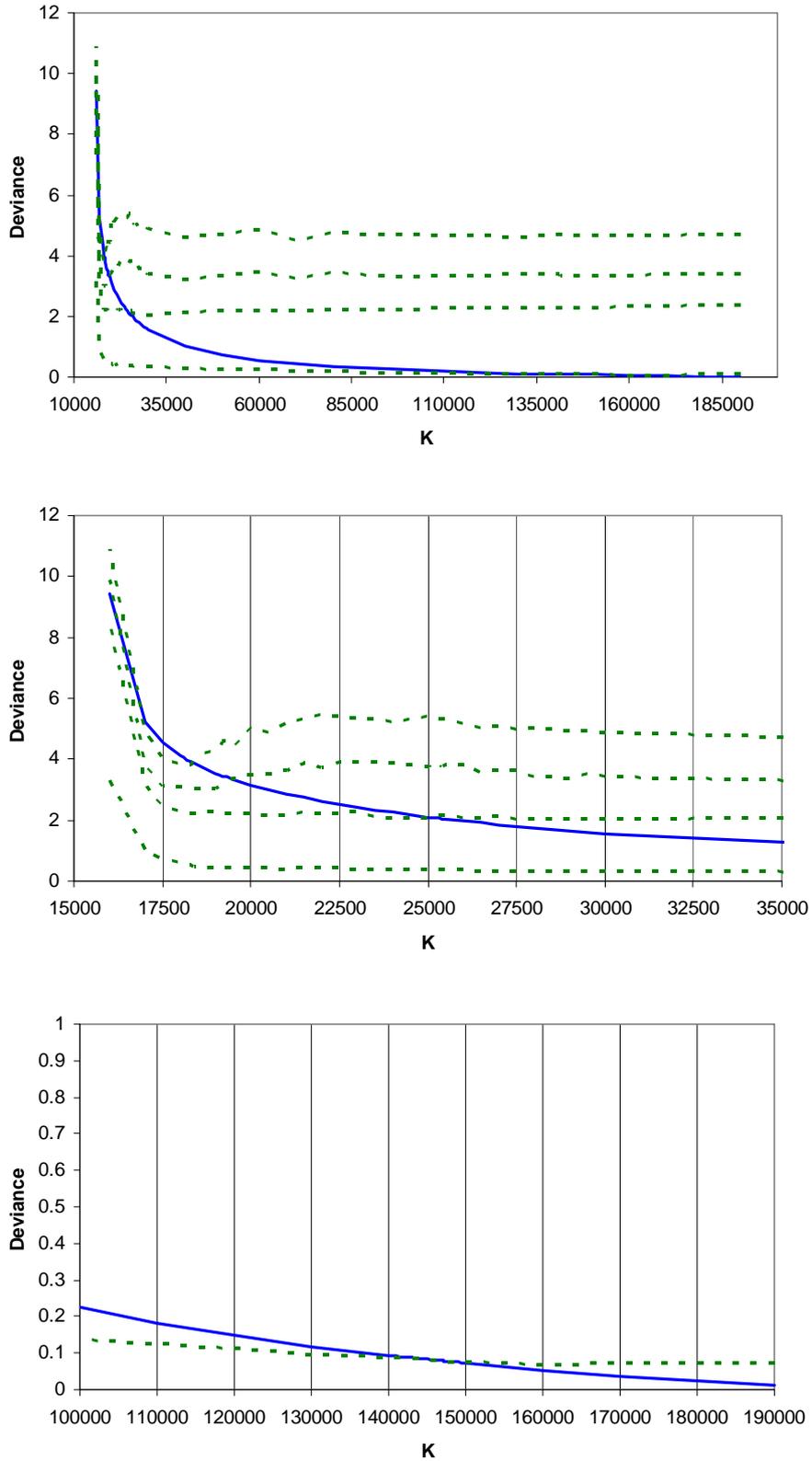
Figure 2. Observed deviance (solid line) and 2.5%, 5%, 10% and 50% quantiles (dashed lines) based on 1000 simulations, together with close ups of the intersections of the quantiles and the observed deviance curve, for **Model 4b** with MSYR = 2%.
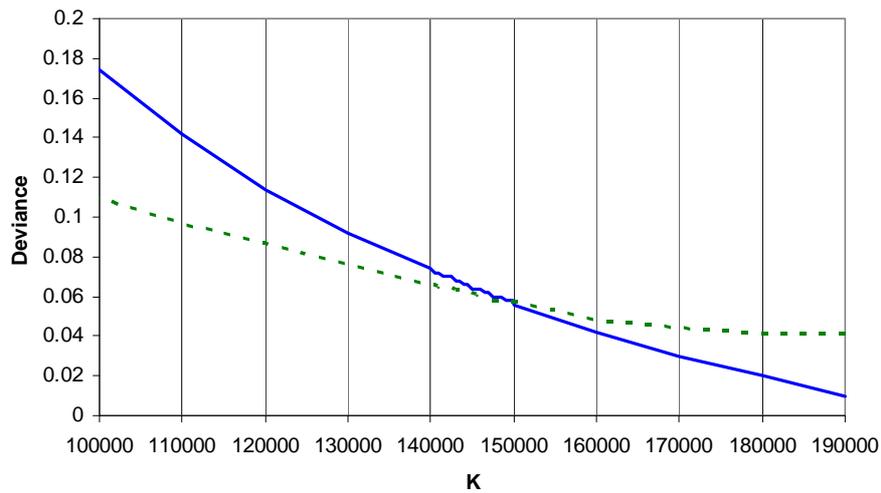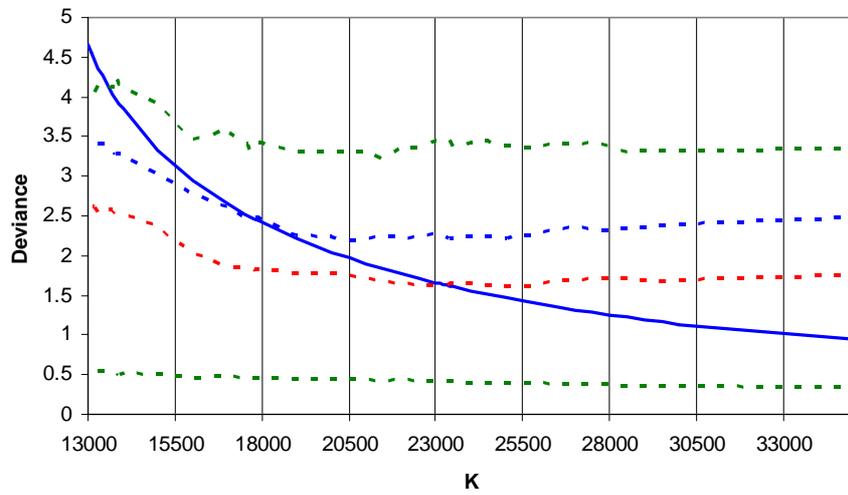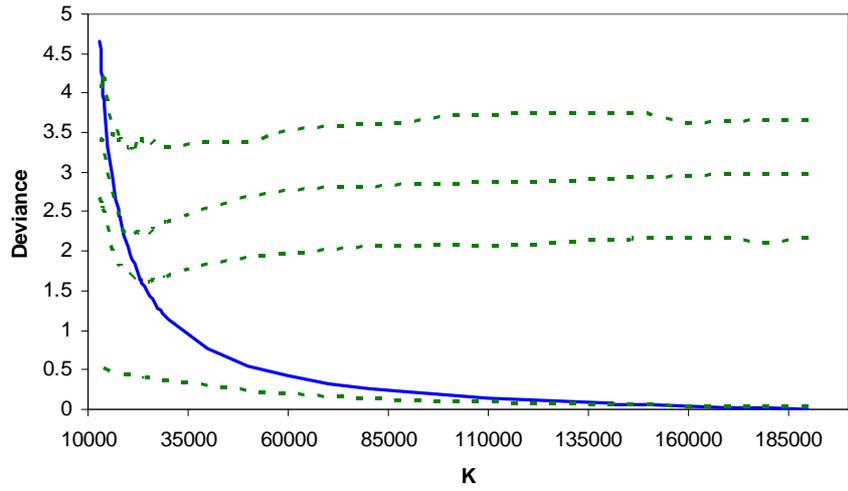
Figure 3. Observed deviance (solid line) and 2.5%, 5%, 10% and 50% quantiles (dashed lines) based on 1000 simulations, together with close ups of the intersections of the quantiles and the observed deviance curve, for **Model 5 (NW+CW)** with MSYR = 2%.

APPENDIX

Simulation algorithm

For the best estimate of virgin biomass (K) (here taken to be 200 000 as a surrogate for infinity), the models described in the text are fitted to the original data to obtain estimates for the overdispersion (σs) and the selectivity of males relative to females (λs) parameters for the period and whalers concerned. Then for a given value of the true virgin biomass (K), and the overdispersion as estimated for K = 200 000, the models are fitted to the original data to obtain estimates of the λs. For each model the deviance as a function of K is obtained for the original data.

For each set of values of K, r, and λ_i, the total annual catches ($C_y = C_y^m + C_y^f$) and the annual reported total catches (C_y^i), the following steps are taken:

Set $N_{1948,a}^g = \frac{1}{2}\eta N_a^*$, using equation (5).

Generate C_y^f, C_y^m, $C_y^{f(i)}$ and $C_y^{m(i)}$.

From $C_{y,a}^g$, project $N_{y,a}^f$ and $N_{y,a}^m$ forward one year (using equations (1) to (3)).

Repeat steps (2) and (3) until the end of the time period (i.e. 2007).

Fit models to the generated data in which K is fixed and for the case when K is estimated to get the deviance value for the generated data.

Repeat steps (1) to (5) 1000 times to get the distribution of the deviance values.

Data generation

The data generation has to take into account that not all whales are sampled for sex, and that there is a period over which both Norwegian and Greenlandic catches occurred. The assumption has been made that the Norwegian catch was always fully sampled, so that the sampled Greenland catch has to be generated from the total Greenland catch each year.

Period 1948–1954 and 1986 (no sampling):

Generate C_y^f from the normal distribution given by:

$$N\left(\frac{\sum_a N_{y,a}^f.G_a}{\sum_a N_{y,a}^f.G_a + \lambda^I \sum_a N_{y,a}^m.G_a}C_y, \sigma_I^2 \frac{\sum_a N_{y,a}^f.G_a}{\sum_a N_{y,a}^f.G_a + \lambda^I \sum_a N_{y,a}^m.G_a}C_y\right),$$
(A.1)

i.e. the λ_i and σ_i correspond to the Greenlandic (1955-1978) period.

The total number of males is then given by $C_y^m = C_y - C_y^f$.

Period 1955–1967 (only Greenland catch, which is sampled):

Generate C_y^f and C_y^m as in 1] above.

Sample C_y^I without replacement and with autocorrelation ρ from C_y with sex split given by C_y^f and C_y^m, to get sampled numbers $C_y^{f(I)}$ and $C_y^{m(I)}$.

Period 1968–1978 (both Greenland and Norwegian catches, both sampled):

Generate the Norwegian catch $C_y^{f(II)}$ from the normal distribution given by:

$$N\left(\frac{\sum_a N_{y,a}^f.G_a}{\sum_a N_{y,a}^f.G_a + \lambda^{II}\sum_a N_{y,a}^m.G_a}C_y^{II}, \sigma_{II}^2 \frac{\sum_a N_{y,a}^f.G_a}{\sum_a N_{y,a}^f.G_a + \lambda^{II}\sum_a N_{y,a}^m.G_a}C_y^{II}\right)$$

The total number of males caught by the Norwegians is then given by $C_y^{m(II)} = C_y^{II} - C_y^{f(II)}$.

Note that the Greenland catch is $C_y - C_y^{II}$, to be comprised of $C_y^{f(I)*}$ females and $C_y^{m(I)*}$ males.

Generate $C_y^{f(I)*}$ from the normal distribution given by:

$$N\left(\frac{\sum_a N_{y,a}^f.G_a}{\sum_a N_{y,a}^f.G_a + \lambda^{I}\sum_a N_{y,a}^m.G_a}\left(C_y - C_y^{II}\right), \sigma_{I}^2 \frac{\sum_a N_{y,a}^f.G_a}{\sum_a N_{y,a}^f.G_a + \lambda^{I}\sum_a N_{y,a}^m.G_a}\left(C_y - C_y^{II}\right)\right)$$

The total number of males caught by Greenland is then given by $C_y^{m(I)*} = \left(C_y - C_y^{II}\right) - C_y^{f(I)*}$.

Sample without replacement and with autocorrelation from $C_y - C_y^{II}$ with sex split given by $C_y^{f(I)*}$ and $C_y^{m(I)*}$, to give the whales caught and sampled by Greenland $C_y^{f(I)}$ and $C_y^{m(I)}$.

Add the $C_y^{f(I)*}$ and $C_y^{m(I)*}$ to the Norwegian generated catches to get the total catches by sex (e.g. $C_y^f = C_y^{f(II)} + C_y^{f(I)*}$).

Period 1979–1985 (both Greenland and Norwegian catches; the former is not sampled, but is assumed to be governed by the parameters for the first (1955–1978) period of sampled Greenland catches):

Generate the Norwegian catch $C_y^{f(II)}$ from the normal distribution given by:

$$N\left(\frac{\sum_a N_{y,a}^f.G_a}{\sum_a N_{y,a}^f.G_a + \lambda^{II}\sum_a N_{y,a}^m.G_a}C_y^{II}, \sigma_{II}^2 \frac{\sum_a N_{y,a}^f.G_a}{\sum_a N_{y,a}^f.G_a + \lambda^{II}\sum_a N_{y,a}^m.G_a}C_y^{II}\right)$$

The total number of males caught by the Norwegians is then given by $C_y^{m(II)} = C_y^{II} - C_y^{f(II)}$.

Note that the Greenland catch is $C_y - C_y^{II}$, to be comprised of $C_y^{f(I)*}$ females and $C_y^{m(I)*}$ males.

Generate $C_y^{f(I)*}$ from the normal distribution given by:

$$N\left(\frac{\sum_a N_{y,a}^f.G_a}{\sum_a N_{y,a}^f.G_a + \lambda^{I}\sum_a N_{y,a}^m.G_a}\left(C_y - C_y^{II}\right), \sigma_{I}^2 \frac{\sum_a N_{y,a}^f.G_a}{\sum_a N_{y,a}^f.G_a + \lambda^{I}\sum_a N_{y,a}^m.G_a}\left(C_y - C_y^{II}\right)\right)$$

The total number of males caught by Greenland is then given by $C_y^{m(I)*} = \left(C_y - C_y^{II}\right) - C_y^{f(I)*}$.

Add the $C_y^{f(I)*}$ and $C_y^{m(I)*}$ to the Norwegian generated catches to get the total catches by sex (e.g. $C_y^f = C_y^{f(II)} + C_y^{f(I)*}$).

Period 1987–2007 (only Greenland catch, which is sampled):

Generate C_y^f from normal distribution given by:

$$N\left(\frac{\sum_a N_{y,a}^f.G_a}{\sum_a N_{y,a}^f.G_a + \lambda^{III}\sum_a N_{y,a}^m.G_a}C_y, \sigma_{III}^2 \frac{\sum_a N_{y,a}^f.G_a}{\sum_a N_{y,a}^f.G_a + \lambda^{III}\sum_a N_{y,a}^m.G_a}C_y\right).$$

The total number of males is then given by $C_y^m = C_y - C_y^f$.

Sample C_y^{III} randomly without replacement from C_y with sex split given by C_y^f and C_y^m, to get sampled numbers $C_y^{f(III)}$ and $C_y^{m(III)}$.

In the data generation algorithm described above, in instances in which a negative catch was generated for one of the sexes, the catch for that sex was set to zero and consequently the catch for the opposite sex was set to the total number being sampled (as otherwise in this case, a catch greater than the number being sampled would have been generated to compensate for the negative generated catch).

Modifications for new applications of this paper
The data generation algorithm described above applies to Model 5, where the total catches and the reported catches refer to those in the NW+CW strata. All the Norwegian catches are assumed to come from the NW+CW population. The catches in the earlier Greenlandic period are split by area by taking males and females separately, calculating the average ratio of the total Greenlandic catch of NW+CW to SW over the years of the later Greenlandic period, and then applying this ratio to the earlier period for which this split is not known. In other respects, the specification of Model 5 follows the prescription given in IWC (2008a).

For the other models considered in this paper, the following alterations to the data generation algorithm were made:

Model 4b: In Step 3 and in Step 4, the Norwegian catches are not generated but the observed Norwegian data is used (assuming that all Norwegian catches were sampled). Therefore the following changes are made:

$$
\begin{aligned}
C_y^{II} &\to C_y^{Norwegian} \\
C_y^{f(II)} &\to C_y^{f(Norwegian)} \\
C_y^{m(II)} &\to C_y^{m(Norwegian)}
\end{aligned}
$$

In Step 5, the total Greenland catches (C_y) have to be split into the NW+CW and SW strata (C_y^{NW+CW} and C_y^{SW}), where these are taken to be the observed data as used for Model 5(NW+CW) and Model 5 (SW). Step 5 is thus changed into:

Generate $C_y^{f(NW+CW)}$ from normal distribution given by:

$$
N\left(\frac{\sum_a N_{y,a}^f.G_a}{\sum_a N_{y,a}^f.G_a + \lambda^{II(NW+CW)}\sum_a N_{y,a}^m.G_a} C_y^{NW+CW}, \sigma_{II(NW+CW)}^2 \frac{\sum_a N_{y,a}^f.G_a}{\sum_a N_{y,a}^f.G_a + \lambda^{II(NW+CW)}\sum_a N_{y,a}^m.G_a} C_y^{NW+CW} \right)
$$

where

- The total number of males is then given by $C_y^{m(NW+CW)} = C_y^{NW+CW} - C_y^{f(NW+CW)}$.

- Sample $C_y^{II(NW+CW)}$ randomly without replacement from C_y^{NW+CW} with sex split given by $C_y^{f(NW+CW)}$ and $C_y^{m(NW+CW)}$, to get sampled numbers $C_y^{f(II(NW+CW))}$ and $C_y^{m(II(NW+CW))}$.

- Generate $C_y^{f(SW)}$ from normal distribution given by:

$$
N\left(\frac{\sum_a N_{y,a}^f.G_a}{\sum_a N_{y,a}^f.G_a + \lambda^{III(SW)}\sum_a N_{y,a}^m.G_a} C_y^{SW}, \sigma_{III(SW)}^2 \frac{\sum_a N_{y,a}^f.G_a}{\sum_a N_{y,a}^f.G_a + \lambda^{III(SW)}\sum_a N_{y,a}^m.G_a} C_y^{SW} \right), \text{ where}
$$

$$
\lambda^{III(SW)} = \lambda_{87}(2006-y)/19 + \lambda_{06}(y-1987)/19,
$$

i.e. λ is assumed to change linearly over time during this period.

- The total number of males is then given by $C_y^{m(SW)} = C_y^{SW} - C_y^{f(SW)}$.

- Sample $C_y^{II(SW)}$ randomly without replacement from with sex split given by $C_y^{f(SW)}$ and $C_y^{m(SW)}$, to get sampled numbers $C_y^{f(III(SW))}$ and $C_y^{m(III(SW))}$.

- Get total catches by sex by adding the catches by sex from each strata (e.g. $C_y^f = C_y^{f(II(NW+CW))} + C_y^{f(III(SW))}$

Model 3:   The data generation algorithm remains essentially the same as for Model 4b, but with the following changes:

- When generating data for the NW+CW strata:

$$\sum_a N_{y,a}^f . G_a \rightarrow \sum_a N_{y,a}^f . G_a \left(A^f \right)$$
$$\sum_a N_{y,a}^m . G_a \rightarrow \sum_a N_{y,a}^m . G_a \left(A^m \right) \text{, where}$$

$$A^f = \frac{e^{\left(\alpha^f + \beta^f y \right)}}{1 - e^{\left(\alpha^f + \beta^f y \right)}}$$
$$A^m = \frac{e^{\left(\alpha^m + \beta^m y \right)}}{1 - e^{\left(\alpha^m + \beta^m y \right)}}$$

- When generating data for the SW strata:

$$\sum_a N_{y,a}^f . G_a \rightarrow \sum_a N_{y,a}^f . G_a \left(1 - A^f \right)$$
$$\sum_a N_{y,a}^m . G_a \rightarrow \sum_a N_{y,a}^m . G_a \left(1 - A^m \right)$$

C_y^{SW}