SC/66a/RMP/5 Rev1

A new variance estimator for Northeast Atlantic minke whales applied to survey data from 1996-2001.

Hans J. Skaug, Hiroko K. Solvang

INTERNATIONAL
WHALING COMMISSION

SC/66/RMP

A new variance estimator for Northeast Atlantic minke whales applied to survey data from 1996-2001

HANS J. SKAUG, HIROKO KATO SOLVANG

Institute of Marine Research, Box 1870 Nordnes, N-5817 Bergen, Norway.
Email: skaug@imr.no

ABSTRACT

The Markov modulated Poisson process is used to estimate variance in whale counts on individual transect legs. This model accounts for overdispersion relative to the Poisson distribution, and constitutes an operationally simpler alternative to the Neyman-Scott process that has been used in the past for Northeast Atlantic minke whales. A second change in methodology is that we suggest replacing the parametric bootstrap method with a somewhat cruder “delta-method” for calculating the variance of the line transect abundance estimator. The new approach is validated on the 1996-2001 surveys and gives CV of 9,3% for the total abundance estimate, while the previous method gave a CV of 10.1%. For individual survey blocks the discrepancy is larger, and in particular the direct measure of overdispersion varies substantially between the old and new method.

MINKE WHALES, NORTH ATLANTIC, ABUNDANCE ESTIMATE, VESSEL SURVEY

INTRODUCTION

As part of the implementation review for Norwegian North Atlantic minke whales at last year’s meeting a new variance estimation approach was discussed (IWC 2014). The purpose of the current manuscript is to describe this method, and to apply it to the 1996-2001 surveys for validation purposes.

The variance of the abundance estimate has traditionally been obtained using a parametric bootstrap approach for this population (Schweder, Skaug et al. 1997, Skaug, Oien et al. 2004, Bøthun, Skaug et al. 2009). This has accounted for uncertainty in both effective strip half width and encounter rate. In particular, it has model overdispersion in encounter rate arising from spatial clustering of whales via the Neyman-Scott process. Although the approach is conceptually appealing, it is computationally demanding and the simulation model is complex due to its flexible nature. The current manuscript presents a simpler approach based on the more traditional “delta method” for combining the two types of uncertainty (strip width and encounter rate). The degree of animal clustering is estimated using the MMPP(Markov modulated Poisson process) which has previously been applied to line transect data (Skaug 2006). We describe how to extend the method Skaug (2006) to independent observer data.

We describe the details of the new approach and apply it to the 1996-2001 surveys for the purpose of comparison.

MATERIAL AND METHODS

A simple design based estimator is used (for each survey block):

$$\hat{N} = \frac{n_A + n_B}{2(\hat{w}_A + \hat{w}_B)L} Area \quad (1)$$

,

where n_A and n_B are the total number of sighted whales from platforms A and B, $\hat{w}_A$ and $\hat{w}_B$ are average estimated effective strip half widths, L is the realized transect length, and *Area* is the area of the survey block. This estimator has been used for all Norwegian surveys (Schweder, Skaug et al. 1997, Skaug, Oien et al. 2004, Bøthun, Skaug et al. 2009). An advantage is that it avoids duplicate identification via use of the term $n_A + n_B$. In the suggested approach, estimation of $\hat{w}_A + \hat{w}_B$ and its standard deviation under a hazard probability model is identical to the earlier approach (Skaug, Øien et al. 2004). The variance of $n_A + n_B$ is estimated for each survey block under the assumption that detected whales are distributed along the transect line according to a MMPP (Skaug 2006). Appendix A adapts this method to independent observers.

The delta method for estimating variance

The purpose of this section is to calculate the variance of (1). In the standard (single observer) line transect literature it is assumed that n and $\hat{w}$ are statistically independent, and thus it is straight forward to find the variance of $\hat{N}$ using the delta method (Buckland, Anderson et al. 1993). The fact that (1) involves double platform (A and B) data does not pose any serious difficulty. However, the fact that $\hat{w}_A + \hat{w}_B$ is not independent across survey blocks, due to pooling of data when fitting the hazard probability model, means that inter-block correlation must be taken into account for total abundance (sum $\hat{N}$ over survey blocks). These aspects can automatically be taken care of by the use of the software package ADMB (ADMB-Project 2009). We now explain how ADMB can be used to performs the delta method to find the variance of the estimator (1). The procedure is not as “clean” as one would like, but having ADMB taking care of all the logistics of the delta method is a great advantage.

Let θ denote the vector of parameters in the hazard probability model, and write $w_A(\theta) + w_B(\theta)$ to emphasize the dependence of w on θ. We let $l_{HP}(\theta)$ denote the log-likelihood function under the hazard probability (HP) model (Skaug, Oien et al. 2004). As a model for encounter rate we assume that $n_A + n_B$ follow a negative binomial distribution (within each survey block), parameterized in terms of the mean $E(n_A + n_B)$ and the overdispersion $\tau = Var(n_A + n_B) / E(n_A + n_B)$. If we denote by D average whale density we have $E(n_A + n_B) = 2(w_A + w_B)L \cdot D$. We define $\delta = (\log(D_1), \ldots, \log(D_J))$ the vector of log densities for each of the J survey blocks, and define a corresponding log-likelihood component $l_{NB}(\delta, \theta; \tau)$ based on the vector (over survey blocks) of $n_A + n_B$. The joint likelihood, which allows ADMB to apply the delta method to (1) is given as

$$l_{HP}(\theta)+l_{NB}(\delta,\theta;\tau) \tag{2}$$

In this likelihood τ is held fixed at a value obtained as explained in Appendix A. When ADMB has maximized (2) to obtain maximum likelihood estimators $(\hat{\theta},\hat{\delta})$ we can calculate the abundance estimate for the j'th survey block

$$\hat{N}_j = \exp(\hat{\delta}_j) Area_j \tag{3}$$

ADMB automatically calculates the standard deviation of (3), and we use this as an approximation to the standard deviation of (1). Figure 1 shows that (1) and (3) give very close results.

Summary of algorithm

1. Estimate $w_A(\theta)+w_B(\theta)$ based on $l_{HP}(\theta)$ using the standard method (Skaug, Oien et al. 2004).
2. Given the output from 1), estimate $Var(n_A+n_B)$ based on the MMPP and Appendix A.
3. Given the estimate of $Var(n_A+n_B)$ from 2) use the ADMB estimate of standard deviation (3).

RESULTS

The method was applied to the survey data from the 1996-2006, and comparison was made to the previous method of (Skaug, Oien et al. 2004). The same set of covariates that was selected as the "best model" in Skaug, Oien et al. (2004) was used. Table 1 show there is relatively good agreement between standard deviations ($SD(\hat{N})$) produced by the new method (MMPP) and previous method (parametric bootstrap). The standard deviation for the total survey area is 10,651 with the new method, while it was 10,821 with the previous method (Skaug, Oien et al. 2004). It should be noted that (for technical reasons) the new model the survey block LOC has been included twice, but LOC is a very small survey block constituting less than 1% of the total abundance.

DISCUSSION

The construction of the likelihood function (2) needed for making ADMB do the delta-method calculations is admittedly somewhat ad-hoc. For instance, it is assumed that n_A+n_B follows a negative binomial distribution which contradicts the assumptions of the MMPP. The only motivation for using the negative binomial distribution is that 1) it is a discrete distribution and 2) the overdispersion can be controlled.

The old and new methods differ in their estimate of $\tau = Var(n_A+n_B)/E(n_A+n_B)$ (Table 1), but there is not a systematic directions in the differences across survey blocks. The differences are particularly striking for the survey blocks NS and LOC. The reasons for the differences are entirely understood.

Appendix A

MMPP based estimation of overdispersion in counts

Skaug (2006) described how to use the Markov modulated Poisson process (MMPP) to modelled clustered line transect data. Only a single observer platform was considered, and here we consider two platforms and show how to calculate the variance of $n_A + n_B$.

Consider the combined platform $A \cup B$. For each observation made by $A \cup B$ there are three possibilities for who actually saw the whale: $u \in \{A, B, AB\}$ (only A, only B, both A and B). The probability of each of these are

$$P(u) = w_{A\cup B}{}^{-1} \begin{cases} w_{A\cup B} - w_B & u = A \\ w_{A\cup B} - w_A & u = B \\ w_A + w_B - w_{A\cup B} & u = AB \end{cases}$$

which follows directly from (3) in Skaug and Schweder (1999). The method in this appendix assumes that estimates of w_A, w_B and $w_{A\cup B}$ are available (from the hazard probability model). Although Skaug (2006) allowed w to vary along the transect line, we shall for simplicity assume that w_A, w_B and $w_{A\cup B}$ are constant within survey block, with values taken to be time averages (taken from the hazard probability model).

Consider in the below arguments a single survey block. Let $n_{A\cup B}$ denote the total number of unique whales seen, and let u be the duplicate identification status: A, B or AB (both A and B). It follows that

$$n_A + n_B = n_{A\cup B} + \sum_{i=1}^{n_{A\cup B}} 1_{u_i = AB} \, ,$$

where $I_{u=AB}$ is an indicator function (1 if seen by both A and B, zero otherwise). By independence of the u_i (conditionally on $n_{A\cup B}$) we have

$$E\left(n_A + n_B \mid n_{A\cup B}\right) = n_{A\cup B}\left\{1 + P(u = AB)\right\},$$
$$Var\left(n_A + n_B \mid n_{A\cup B}\right) = n_{A\cup B} P(u = AB)\left\{1 - P(u = AB)\right\}.$$

The law of total variance gives

$$Var\left(n_A + n_B\right) = E\left\{Var\left(n_A + n_B \mid n_{A\cup B}\right)\right\} + Var\left\{E\left(n_A + n_B \mid n_{A\cup B}\right)\right\}$$
$$= Var(n_{A\cup B})\left\{1 + P(u = AB)\right\}^2 + E(n_{A\cup B}) P(u = AB)\left\{1 - P(u = AB)\right\}.$$

Formulas for $E(n_{A\cup B})$ and $Var(n_{A\cup B})$ under the MMPP can be found in eqn. (7) and (8) of Skaug (2006)

REFERENCES

ADMB-Project (2009). AD Model Builder: automatic differentiation model builder. Developed by David Fournier and freely available from admb-project.org.

Buckland, S. T., et al. (1993). Distance Sampling, London: Chapman\&Hall.

Bøthun, G., et al. (2009). Abundance of minke whales in the Northeast Atlantic based on survey data collected over the period 2002-2007.

IWC (2014). "Report of the Scientific Committee. Annex D. Report of the Sub-Committee on the Revised Management Procedure." {Journal of Cetacean Research and Management (Suppl.)}.

Schweder, T., et al. (1997). "Abundance estimates for northeastern Atlantic minke whales. Estimates for 1989 and 1995." {Rep. int. Whal. Commn.} **47**: 453-484.

Skaug, H. J. (2006). "Markov modulated poisson processes for clustered line transect data." Environmental and Ecological Statistics **13**(2): 199-211.

Skaug, H. J., et al. (2004). "Abundance of minke whales (Balaenoptera acutorostrata) in the Northeast Atlantic: variability in time and space." Canadian Journal of Fisheries and Aquatic Sciences **61**(6): 870-886.

Skaug, H. J. and T. Schweder (1999). "Hazard models for line transect surveys with independent observers." Biometrics **55**: 29--36.

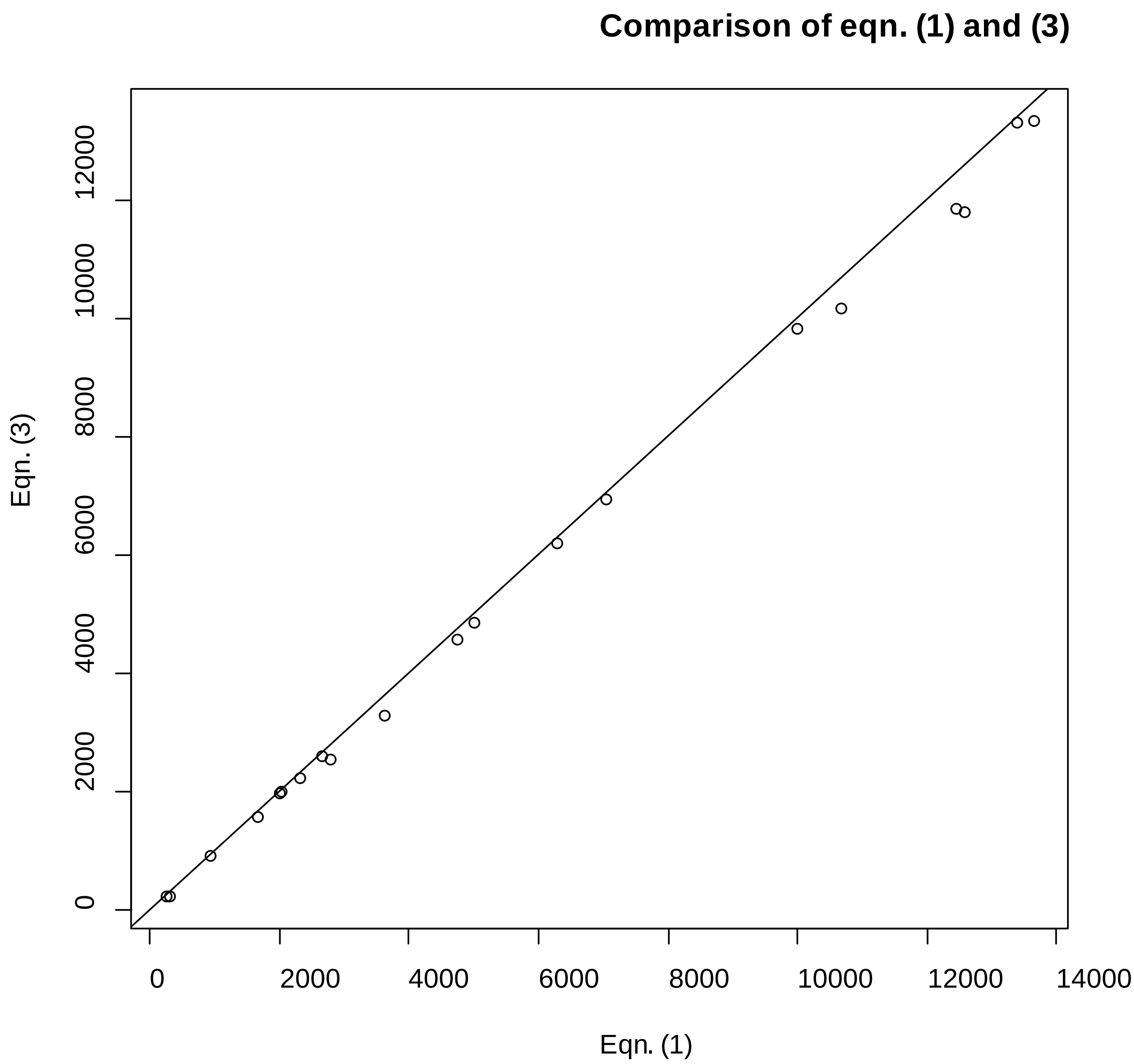
Skaug, H. J., et al. (2004). "Current abundance of minke whales (Balaenoptera acutorostrata) in the Northeast Atlantic; variability in time and space." Canadian Journal of Fisheries and Aquatic Sciences **61**: 870-886.

.

Table 1 Comparison of MMPP to the previous parametric bootstrap approach. SMA is small area, V is variance, τ is over dispersion, PB is the parametric bootstrap approach of (Skaug, Oien et al. 2004).

			Effective strip half width					τ		SD($\hat{N}$)	
SMA	Block	Year	n_A+n_B	w_A	w_B	w_{AUB}	V(n_A+n_B)	MMPP	PB	MMPP	PB
CM	JMC	1997	38	241	214	370	125	4	2	1483	921.23
	NVN	1997	59	274	227	401	121	3	2	2300	1789.07
	NVS	1997	84	292	249	426	357	5	4	3263	3426.41
EB	BAE	2000	73	243	210	368	564	9	11	4133	4887.62
	FI	1996	120	267	222	396	419	4	6	1287	1563.04
	GA	2001	56	229	190	344	363	7	8	3582	3730.05
	KO	2001	25	288	249	426	130	5	3	1192	819.23
	NOS	1996	125	236	209	362	430	3	4	2155	2477.94
EC	LOC96	1996	2	233	207	359	5	1	10	184	796.69
	LOC00	2000	10	261	219	388	13	1	11	349	860.98
EN	NS	1998	146	214	184	328	2692	21	57	4555	3455.07
	NSC	1998	43	224	193	341	96	2	2	1254	1368.45
ES	BAW	1999	30	366	291	513	295	8	6	1656	1516.23
	BJ	1999	43	481	395	658	91	2	3	475	403.40
	NON	1999	24	227	195	345	35	2	2	683	703.61
	SV	1999	59	312	262	453	323	5	3	1407	1213.78
	SVI	1999	10	230	196	348	8	1	4	726	1314.58
	VSI	1999	25	603	493	798	239	15	7	194	140.33
	VSN	1999	43	308	259	447	183	5	3	540	304.38
	VSS	1999	63	310	269	455	438	3	9	534	859.82

Figure 1


Comparison of eqn. (1) and (3)
Eqn. (3)
0
2000
4000
6000
8000
10000
12000
0
2000
4000
6000
8000
10000
12000
14000
Eqn. (1)