

Quantifying Some of the Major Sources of Uncertainty Associated with Estimates of Harp Seal Prey Consumption

Part I: Uncertainty in the Estimates of Harp Seal Population Size

William G. Warren, Peter A. Shelton and Garry B. Stenson

Science Branch, Department of Fisheries and Oceans, Northwest Atlantic Fisheries Centre
P. O. Box 5667, St John's, Newfoundland, Canada A1C 5X1

Abstract

An effort is made to quantify some of the major sources of uncertainty associated with estimates of harp seal (*Phoca groenlandica*) prey consumption, with special reference to the uncertainty in estimates of seal population size. A population dynamics model is used, the inputs for which are the annual catches at age and pregnancy rates. Two parameters, instantaneous mortality rate, and the hunting selection on the pups for years prior to the availability of pregnancy data, are estimated by non-linear least squares using available survey estimates of pup production. The uncertainty in estimates of population size is estimated by Monte Carlo methods from the estimates of sampling error in the pregnancy rates and in the survey estimates of pup production, with the catches at age assumed to be known. The uncertainty in the population estimates is dominated by the uncertainty in the survey estimates.

Key words: catch rates, harp seal, mortality, pregnancy rates, stock size

Introduction

The harp seal (*Phoca groenlandica*) population in the Northwest Atlantic was estimated to number 4.8 million in 1994 and to be increasing at about 5% per year (Shelton *et al.*, 1996). There is considerable interest in what impact predation by this pinniped might have on fish populations. Prey consumption depends not simply on the size of the population but also, *inter alia*, on the seasonal and spatial distribution of the predator, its energy requirements, the calorific content of the various prey species, and the proportion of these species in the diet. There is uncertainty in our knowledge of all these components which, of course, carries through to uncertainty in any estimate of prey consumption. Our information on these inputs ranges from sample estimates, sometimes with conventional measures of precision (standard errors), to guesses based on unquantified observation.

This paper represents an attempt to quantify the effect of these various sources of uncertainty on the estimate of annual prey consumption, in particular, the amounts of Atlantic cod (*Gadus morhua*), Arctic cod (*Boreogadus saida*) and capelin (*Mallotus*

villosus) in NAFO Div. 2J and 3KL (Stenson *et al.*, 1997), and to determine which of the components contributes most to this uncertainty. The paper describes the determination of the accuracy of estimates of the size of the harp seal population through estimates of pup production. The impact of this uncertainty on the estimation of consumption is presented in Shelton *et al.*, 1997.

Methods

A population dynamics model

Our starting point is the population dynamics model given in Cadigan and Shelton (1993) which is based on an earlier model presented by Roff and Bowen (1983) using the cohort analysis equation of Pope (1972). Specifically, the population number at age a in year t , $n_{a,t}$ is given by:

$$n_{a,t} = (n_{a-1,t-1} e^{-m/2} - c_{a-1,t-1}) e^{-m/2}, 0 < a < A$$
$$n_{o,t} = \sum_i n_{i,t} f_{i,t}$$

where $c_{a,t}$ is the number at age a caught in year t , $f_{a,t}$ is the per capita pregnancy (fecundity) rate of

age a parents in year t assuming a 1:1 sex ratio and m is the instantaneous mortality rate, assumed to be constant, i.e. independent of t , etc. A plus age-class, A , i.e. ages A and greater is included in the model so that:

$$n_{A,t} = (n_{A-1,t-1} e^{-m/2} - c_{A-1,t-1}) e^{-m/2}$$

where $A-1$ is to be read as all ages $A-1$ and greater.

To accommodate numbers at age for years prior to the first year for which pregnancy data are available, some modifications to the model are required. Let t_o denote the first year for which pregnancy data are available. Further, assuming that seals do not live longer than A_m years, thus obviates the need for a plus age-class. Under the assumption that pup production prior to t_o is equal to a hunting selection parameter, s , times the pup production, where, as with the natural mortality rate, m , s is taken to be constant, we have:

$$n_{a,t_o-1} = s e^{-ma} c_{o,t_o-a-1} - \sum_{i=1}^a e^{-m(i-1/2)} c_{a-i,t_o-i-1}, \quad a = 1, 2, \dots, A_m$$

Cadigan and Shelton (1993) transform the above to a "statistical" model (Appendix 1) in which the numbers at age, $n_{a,t}$, appear solely on the left-hand side of the system of equations while the right-hand side contains the parameters to be estimated, s and m , and the catches at age, $c_{a,t}$ and fecundities, $f_{a,t}$, which are assumed to be known. With time standardized so that $t_o = 0$ the resulting equations are for pups:

$$n_{0,t} = \sum_{i=1}^{A_m} \left[s e^{-m(t+i-1)} w_t^i(1) - e^{-m(t+i-1/2)} u_t^i(1) \right] - \sum_{i=1}^t e^{-m(t-i+1/2)} v_t^i(1)$$

and for total population numbers:

$$n_{.,t} = \sum_{i=1}^{A_m} \left[s e^{-m(t+i-1)} w_t^i(.) - e^{-m(t+i-1/2)} u_t^i(.) \right] - \sum_{i=1}^t e^{-m(t-i+1/2)} v_t^i(.)$$

The vectors u_t^i , v_t^i and w_t^i are defined in the Appendix; $w_t^i(1)$ is the first, i.e. age zero, element of w_t^i ,

etc. and $(.)$ denotes the sum over all ages. Note that the u_t^i , v_t^i and w_t^i are functions solely of quantities assumed to be known, namely catch numbers and pregnancy rates. Thus, the expression for $n_{0,t}$ has the form $sg_1(m) - g_2(m)$, where $g_1(m)$ and $g_2(m)$ are functions involving m but not s .

Let \hat{n}_{0,t_i} be an estimate of n_{0,t_i} for those years, t_i in which a survey estimate of pup production is available, and let $\hat{\sigma}_{t_i}^2$ be an estimate of its variance. Cadigan and Shelton (1993) then assume that $\hat{n}_{0,t_i} \sim N(n_{0,t_i}, \hat{\sigma}_{t_i}^2)$ thus enabling s and m to be estimated by weighted non-linear least squares in a relatively straightforward manner.

Input Parameters

Survey Estimates of Pup Production. Mark-recapture estimates of pup production are available for 1978–80 and 1983 (Bowen and Sergeant, 1983; MS 1985; Roff and Bowen, 1986). Based on the critical review of these estimates by Warren (MS 1991), who also took into account the Cooke *et al.* (1985) review of harp-seal population dynamics, the estimate that appeared to be the most viable for each of the four years was used for parameter estimation. In addition, pup production in 1990 and 1994 was estimated based on regularly spaced photographic and visual aerial surveys of the whelping patches (Stenson *et al.*, 1993; 1995). Because sampling was systematic, the sampling errors associated with the surveys were estimated from the serial differences. Additional components of variance were incorporated to take into account possible errors in the adjustment made for pups not on the ice at the time of the survey (either not yet born or already weaned) and, in the case of the photographic surveys, corrections made to the counts because of reader bias. Details are given in Stenson *et al.*, (1993, 1995). The six estimates, with their estimated standard errors, are included in Table 1.

Pregnancy Rates. Late-term pregnancy rates were obtained by examining reproductive tracts of female harp seals collected off Newfoundland between October and February (Sjare *et al.*, 1996a). Age specific fecundity rates were available by age (3, 4, 5, 6, and 7+) for 1954, 1965–70, 1978–82, 1985–94 (see Sjare *et al.*, 1996a). Sample sizes in the 1980s were small (usually less than 10 for combinations of year and age) and only slightly greater in the 1990s. Indeed, in the 1980s there were two years where one age class is not represented and some combinations of year and age when only one seal was examined. From 1965 to 1970 and in 1978,

TABLE 1. Survey estimates of pup production of harp seals in the Northwest Atlantic. Methods refer to either mark recapture techniques (MR) or visual and/or photographic aerial surveys (AS).

| Year | Method | Estimate | Standard error |
|------|--------|----------|----------------|
| 1978 | MR | 497 000 | 34 000 |
| 1979 | MR | 478 000 | 35 000 |
| 1980 | MR | 475 000 | 47 000 |
| 1983 | MR | 534 000 | 33 000 |
| 1990 | AS | 577 900 | 38 800 |
| 1994 | AS | 702 900 | 63 600 |

sample sizes were usually greater (>15 per year and age combination), although it should be noted that no sampling was undertaken from 1971 to 1977, nor in 1983 and 1984.

In addition to the question of what to do for those years for which pregnancy data are missing, it would seem precarious to use the small-sample data as they stand. For example, with only one seal examined the pregnancy rate would be estimated as either 0 or 100%, neither of which is realistic. With sample sizes of 5 or less, which make up about half the data of the 1980s, there would seem to be an unacceptably high probability of having individual estimates appreciably different from the true rate. Given the lack of data in certain years and the small sample sizes in others, we attempted to determine, with certain provisos, the most parsimonious representation of pregnancy rates that was consistent with the data. To do this, let N_i be the number of seals examined in year i and X_i denote the number of these that are pregnant. Let p_i be the (unknown) pregnancy rate in year i . Then the likelihood of the observations is:

$$\prod_i \frac{N_i!}{X_i!(N_i - X_i)!} p_i^{X_i} (1 - p_i)^{N_i - X_i}$$

It is more convenient to work with the logarithm of the likelihood which, with the terms not involving the p_i omitted, is:

$$\sum_i [X_i \log(p_i) + (N_i - X_i) \log(1 - p_i)]$$

This is maximized by taking the $p_i = X_i / N_i$, which are the maximum likelihood estimates, m.l.e.

The most parsimonious representation possible is for all the p_i to be equal. Under this hypothesis,

the (log) likelihood becomes:

$$\log(p) \sum_i X_i + \log(1 - p) \sum_i (N_i - X_i)$$

and the m.l.e. $p = \sum_i X_i / \sum_i N_i$. Asymptotically, under the hypothesis, twice the negative of the difference between the log likelihoods would be distributed as χ^2 on, here 21 degrees of freedom, d.f. (Since there are 22 years and, under the hypothesis, one parameter is being estimated). For all age classes, the hypothesis of a single pregnancy rate has to be rejected.

It would seem reasonable to suppose that pregnancy rates in successive years would be similar. The following strategy was therefore adopted. First, a 2×2 contingency table was formed from the data of the first two years, namely:

| | | |
|-------------|-------------------------|-------------|
| X_1 | $N_1 - X_1$ | N_1 |
| X_2 | $N_2 - X_2$ | N_2 |
| X_{\cdot} | $N_{\cdot} - X_{\cdot}$ | N_{\cdot} |

where $X_{\cdot} = X_1 + X_2$, etc. The conventional χ^2 statistic was calculated for this table and if the null hypothesis (of common pregnancy rate) was accepted (at the 5% level) these data were pooled and a new 2×2 table formed by including the next year's data, namely:

| | | |
|-------------|-------------------------|-------------|
| X_{\cdot} | $N_{\cdot} - X_{\cdot}$ | N_{\cdot} |
| X_3 | $N_3 - X_3$ | N_3 |
| X_{\cdot} | $N_{\cdot} - X_{\cdot}$ | N_{\cdot} |

where $X_{\cdot} = X_{\cdot} + X_3$, etc. This procedure was continued as long as the successive χ^2 values remained non significant (5% level). When a significant χ^2 value was encountered, the sequence was

terminated and a new sequence begun, starting with the year for which a significant change in pregnancy rate was indicated. Since the χ^2 test is asymptotic and may be unreliable for the smaller sample sizes, it was replaced by Fisher's (1935) exact test.

The sequential χ^2 tests were carried out moving forward in time. They were also carried out moving backward in time from 1994, yielding essentially the same result. Sometimes in moving forward, a change in pregnancy rate would be indicated between two years (e.g. 1987 and 1988), whereas, in moving backward, the change would be placed between two slightly different years (e.g. 1987 and 1986). Such cases were resolved by computing the log likelihood for the two possibilities and the one with the greater likelihood selected.

A priori, it was decided that single years should not be isolated with rates different from the neighbouring years. There were only 3 such instances and in these cases the rates were merged with those of the previous and subsequent years. Finally, on viewing the overall estimates, it was found that the estimates for age 5 would be more consistent with those for the other ages if initial groupings of 1978–82 and 1985–94 were taken as 1978–88 and 1989–94. Although this departed from the sequential procedure as described, the overall likelihood was again improved, albeit slightly.

For ages 6 and 7+, the pregnancy rates appeared to be the same prior to and after the gap in the data from 1971 to 1977. However, for the younger ages, there was a marked increase between 1970 and 1978. For these ages the average of the rates before and after the break was used for the period from 1971 to 1977. With the amendment of the age 5

estimates, there was no indication of a change in rate between 1982 and 1985. The resulting estimates are given in Table 2.

There is strong evidence for a decline in pregnancy rates in recent years, starting with the younger seals in the late-1980s. Whether the drop is as sharp as indicated in Table 2 is debatable, however various approaches at smoothing the transition resulted in only minor differences in the estimates of pup production or total population size.

Catch Data. Catch-at-age data from 1952 to 1993, initially compiled by Sjare *et al.*, (1996b), are given in Shelton *et al.* (1996). Although there is the possibility of under-reporting, of incorrect aging and the misapportioning to individual years those captures reported by broader age classes, these data are, for the purpose of this study, assumed to be correct as reported.

Results and Discussion

Using Asymptotic Properties. Using these data Shelton *et al.* (1996) estimated pup production and total population size in 1994. They consider two formulations, one in which the mortality rate of pups was assumed to be the same as that for all other ages and one in which it was assumed to be three times that of the other ages. The differences in the estimates turn out to be inconsequential; the results presented below are based on the first-mentioned and more parsimonious assumption.

The pup production in 1994 was estimated as 714 525, slightly greater than the survey estimate, and the total population size as 4 759 984 (Shelton *et al.*, 1996). These followed from estimates of s

TABLE 2. Estimated pregnancy rates of harp seals in the Northwest Atlantic obtained from Sjare *et al.* (1996a).

| Year | Age | | | | |
|---------|--------|--------|--------|--------|--------|
| | 3 | 4 | 5 | 6 | 7+ |
| 1954–67 | 0.0172 | 0.1818 | 0.5435 | 0.7321 | 0.8648 |
| 1968–70 | 0.0172 | 0.1818 | 0.5435 | 0.8684 | 0.8648 |
| 1971–77 | 0.0570 | 0.3662 | 0.7162 | 0.8684 | 0.8648 |
| 1978–87 | 0.0968 | 0.5507 | 0.8043 | 0.8684 | 0.8648 |
| 1988 | 0.0968 | 0.1467 | 0.8043 | 0.8684 | 0.8648 |
| 1989 | 0.0615 | 0.1467 | 0.4048 | 0.8684 | 0.8648 |
| 1990–94 | 0.0615 | 0.1467 | 0.4048 | 0.6154 | 0.6341 |

and m of 2.912801 and 0.107364, respectively, i.e. a mortality rate of about 11% and an exploitation rate on pups of $1/2.912801 \approx 34\%$ for years 1952–54.

The asymptotic standard errors of the estimates of s and m were estimated to be 0.084319 and 0.003188, respectively, with correlation 0.999677. High correlation between parameter estimates is common in growth models.

Shelton *et al.* (1996) used these estimates (and their asymptotic normality) to construct a Monte Carlo estimate of the 95% confidence interval for the population size. The limits were estimated as approximately 4.1–5.0 million. It will be noticed that, although symmetric in probability, these limits are asymmetric. The Monte Carlo generated distribution of population size (Fig. 6 in Shelton *et al.*, 1996) exhibited a strong negative skewness. These results were confirmed by an independent simulation, based on 1 000 Monte Carlo realizations, which yielded 95% limits of 4 049 584 – 4 936 532 and the distribution presented in Fig. 1. (There is a slight difference between Fig. 1 and Fig. 6 of Shelton *et al.* (1996); the groupings in Fig. 1 are centred on the 0.05 millions, i.e. ... 4.55, 4.65, 4.75, ... whereas in Shelton *et al.* (1996) they are centred on the 0.1 millions, i.e. ... 4.5, 4.6, 4.7, ...).

Using an Alternative Non-Asymptotic Approach. With only six years of survey data from

which to estimate the parameters, one may question the viability of the asymptotic variances (standard errors) of the estimates and their asymptotic normality. Monte Carlo simulation can, however, be moved back to an earlier stage in the estimation process. For example, we have the estimated standard errors for the survey estimates of pup production; indeed, it was assumed that $\hat{n}_{0,t} \sim n(n_{0,t}, \hat{\sigma}_t^2)$. For the purpose of determining how much the estimates of s and m , and thence the estimate of pup production and population size, would vary if the survey estimates departed from the obtained values, survey estimates may be simulated as $N(\hat{n}_{0,t}, \hat{\sigma}_{t_i}^2)$ random variables.

A similar approach may be applied to the pregnancy rates. The estimates are of the form X/N . Now $\theta = \arcsin \sqrt{(X/N)}$, in radians, is approximately normally distributed with standard deviation $s_\theta = \sqrt{821/N} \times \pi / 180$. Accordingly, simulated pregnancy rates, consistent with the data, may be constructed as $\sin^2(\theta + s_\theta z)$, where z is generated as a standard normal variable.

Simulations were performed by (1) varying the survey estimates of pup production with the pregnancy rates unchanged, and (2) varying both survey estimates of pup production and pregnancy rates. These simulations, particularly (2), require considerably greater computational effort than those using the asymptotic variances of the estimates of

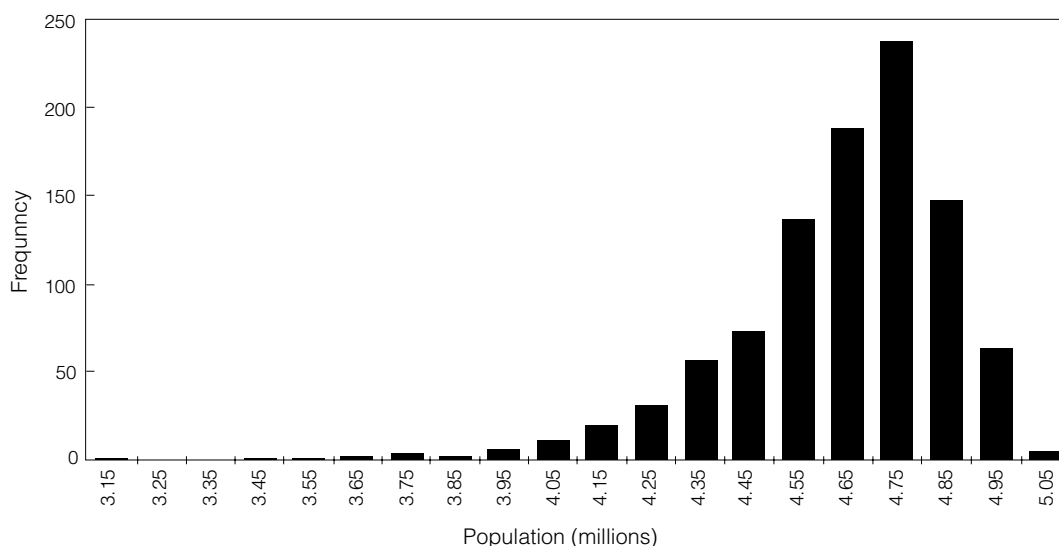


Fig. 1. Histogram of Monte Carlo generated population size based on the asymptotic normality and asymptotic variance-covariance of the estimates of s and m (1 000 realizations).

s and m . For this reason only 100 realizations of each were carried out instead of 1 000. Nevertheless, the results clearly appear to differ from those based on the asymptotics.

We first look at the realized distributions of the estimates of s and m . Summary statistics are given in Table 3.

The means were in excellent agreement with the initial point estimates (2.923700 and 0.107786); the standard deviations, however, were approximately double the asymptotic estimates (0.084319 and 0.003188). That the standard deviations with both survey estimates and pregnancy rates varied should be greater than those with only the survey estimates varied was to be expected, as was the reduction in the correlation. What is, perhaps, surprising, is how small this increase was. Histograms of the generated s and m are given in Fig. 2.

Summary statistics for the population estimates for 1994 are given in Table 4.

There was, again, excellent agreement between the means and the point estimate (4 759 985) and, as would be expected, the standard deviation was somewhat greater when both pregnancy rates and survey estimates were varied. What is, perhaps, surprising, were the somewhat stronger (negative) correlations with s and m . Histograms of the generated n are given in Fig. 3. These give no suggestion of the negative skewness observed using asymptotic properties. The 95% confidence limits using this alternative approach would be placed at approximately 4.10–5.41 million with survey estimates varied, and 4.09–5.53 million with both survey estimates and pregnancy rates varied.

Pup production estimates for 1994 were likewise simulated. These were done independently of

the population numbers although they could have been done simultaneously. The summary statistics are given in Table 5. The mean n_0 were greater, but less than 1% greater, than the point estimate (714 525). The statistics for s and m should be the same as in Table 3 and again, given that these were based on only 100 Monte Carlo realizations, the differences were inconsequential. The histograms (Fig. 4) were comparable with those of Fig. 2. This, in effect, gave us 200 realizations for the distributions of s and m ; the pooled histograms are given in Fig. 5. The one notable feature was the weaker (negative) correlations between n_0 and s or m when both survey estimates and pregnancy rates were varied.

The histograms for the simulated pup production numbers are given in Fig. 6. These, as for the total population (Fig. 3), appear reasonably symmetric. Approximate 95% confidence limits would be 633 000–819 000 when the survey estimates are varied, and 618 000–819 000 when both survey estimates and pregnancy rates are varied.

Discussion

Computational considerations

When the pregnancy rates are varied, each iteration requires a time-consuming calculation to generate the set of non-linear equations from which s and m are estimated. This is followed by their actual estimation by weighted non-linear least squares, which, although not as time consuming as the generation of the set of equations, can take an appreciable amount of time. The latter also applies when just the survey estimates of pup production are varied. However, once s and m are determined, and given the pregnancy rates, the calculation of pup production, total population size and, indeed, numbers-at-age for all years of the study is very rapid. This raises the question of whether one can, instead of generating variable pregnancy rates and

TABLE 3. Parameters of the Monte Carlo generated distribution of the estimates of s and m .

| Parameter | Survey only | Survey and Pregnancy |
|------------------------|-------------|----------------------|
| Mean s | 2.917856 | 2.937480 |
| Mean m | 0.107212 | 0.107532 |
| Standard deviation s | 0.159921 | 0.196255 |
| Standard deviation m | 0.006026 | 0.007291 |
| Correlation | 0.996480 | 0.988190 |

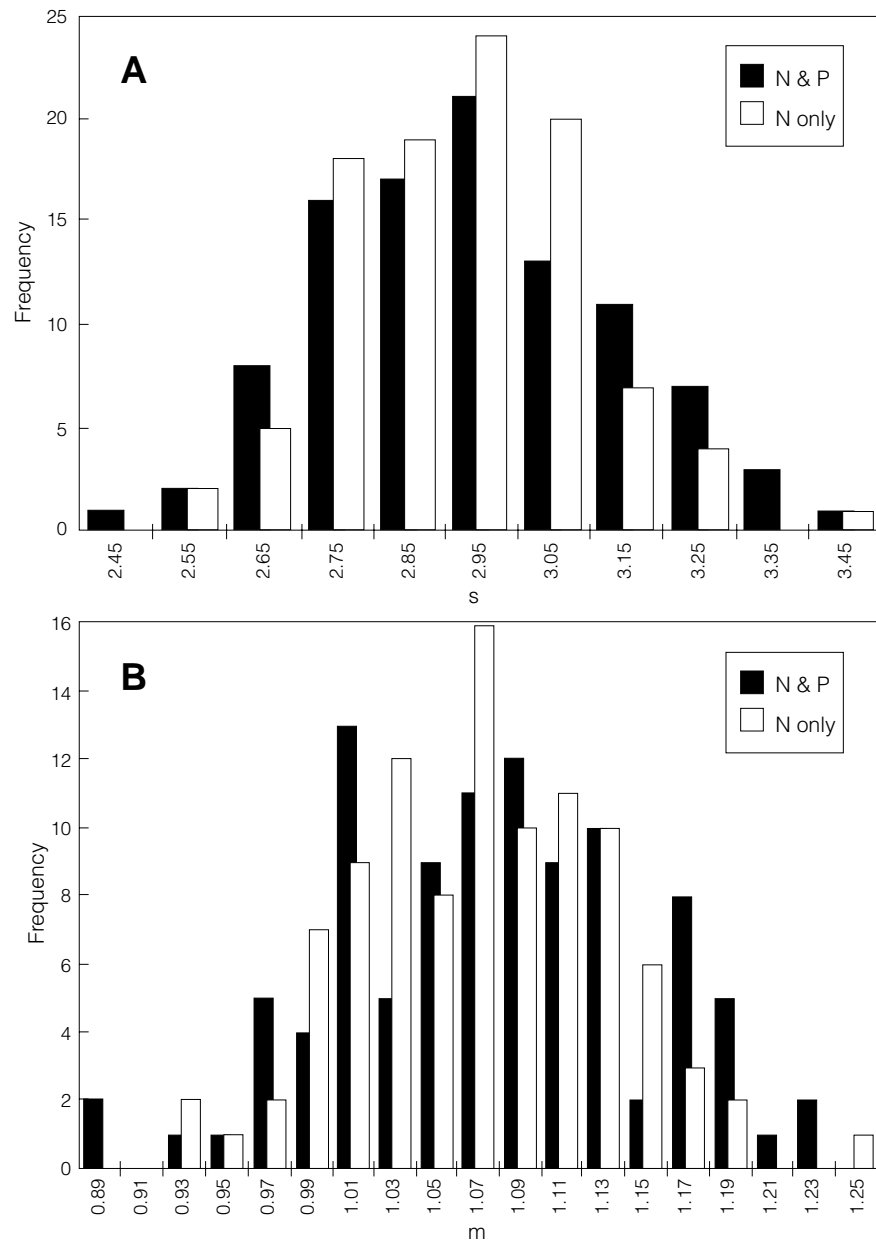


Fig. 2. Histogram of Monte Carlo generated values of (A) s and (B) m obtained when generating total population size; N and P – both survey estimates and pregnancy rates varied, N only – only survey estimates varied (100 realizations).

survey estimates of pup production, generate s and m directly in such a way as to be equivalent to their generation by varying pregnancy rates and survey estimates?

From the simulations, we have the realized values of s and m from which their joint distribution

can be studied. Their means, variances and covariance (correlation), based on 100 realizations, have been presented. Recall that, although simulations were done separately for pup production and total population, for each of these the construction of s and m is identical and we have, in effect, 200 realizations.

TABLE 4. Parameters of the Monte Carlo generated distribution of estimated 1994 population size.

| Parameter | Survey only | Survey and Pregnancy |
|--------------------------|-------------|----------------------|
| Mean n . | 4 745 856 | 4 751 650 |
| Standard deviation n . | 297 794 | 351 451 |
| Correlation n , s | -0.800764 | -0.835915 |
| Correlation n , m | -0.813437 | -0.850516 |

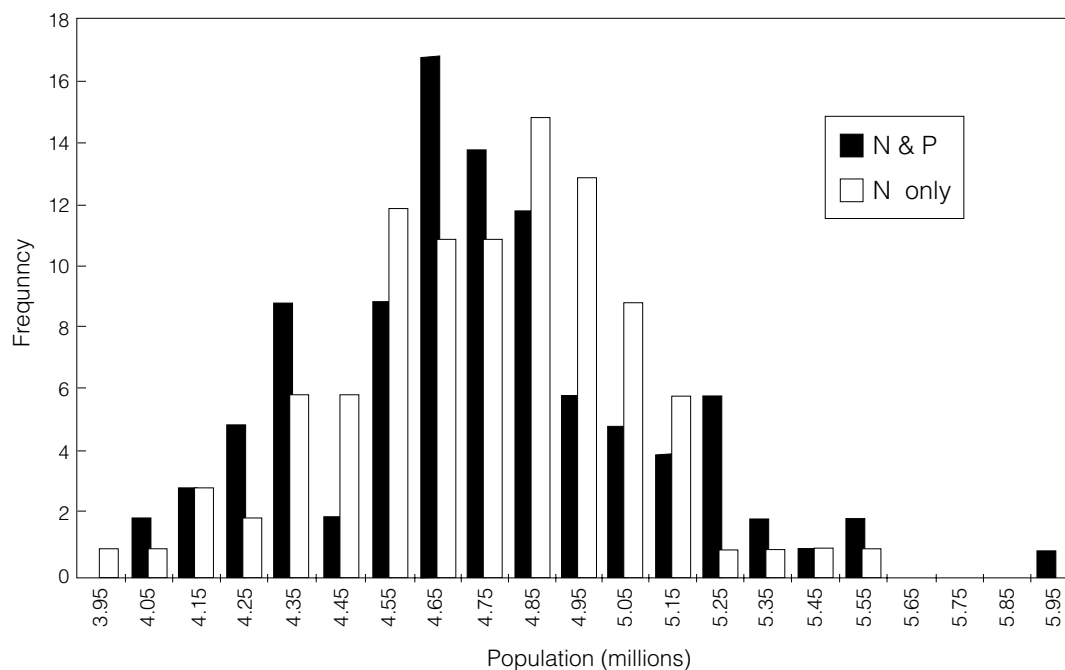


Fig. 3. Histogram of Monte Carlo generated population size; N and P – both survey estimates and pregnancy rates varied, N only – only survey estimates varied (100 realizations).

TABLE 5. Parameter estimates of the Monte Carlo generated distribution of the estimates of s and M and the estimated 1994 pup production.

| Parameter | Survey only | Survey and Pregnancy |
|--------------------------|-------------|----------------------|
| Mean n_0 | 719 961 | 720 743 |
| Mean s | 2.910435 | 2.896325 |
| Mean m | 0.106853 | 0.106228 |
| Standard deviation n_0 | 49 182 | 44 617 |
| Standard deviation s | 0.173824 | 0.175254 |
| Standard deviation m | 0.006428 | 0.007142 |
| Correlation n_0 , s | -0.827514 | -0.674195 |
| Correlation n_0 , m | -0.848347 | -0.687709 |
| Correlation s , m | 0.994853 | 0.985293 |

 n_0 based on 200 realizations.

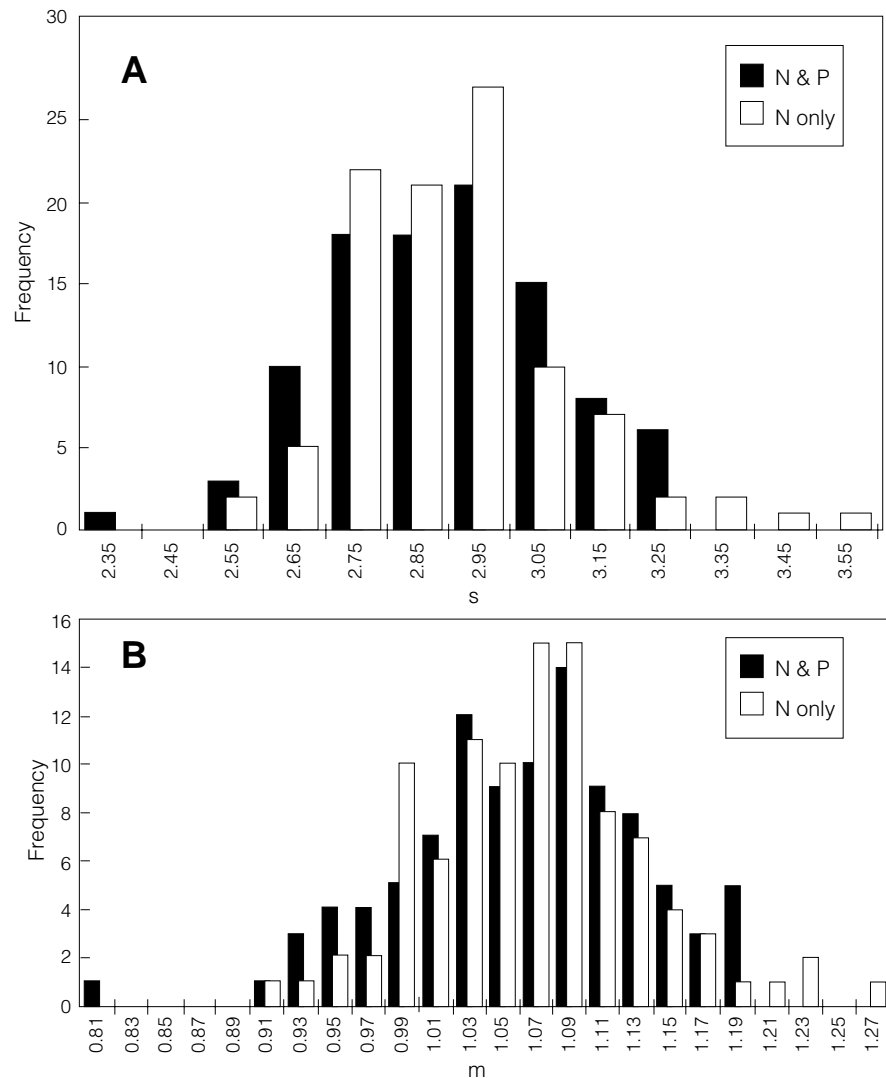


Fig. 4. Histogram of Monte Carlo generated values of (A) s and (B) m obtained when generating pup production; N and P – both survey estimates and pregnancy rates varied, N only – only survey estimates varied (100 realizations).

It would be convenient if the s and m followed a bivariate normal distribution. To test this the Anderson-Darling statistic has been used. Since the mean and variance have to be estimated from the data, the modification, A^* , was calculated; 10%, 5%, 2.5% and 1% critical values are 0.631, 0.752, 0.873, 1.035, respectively (Stephens, 1974, 1982). For s and m we obtained $A^* = 0.311$ and 0.330, respectively. Thus, with an estimated correlation of 0.9857, it would seem that the assumption of bivariate normality is justified.

The above result was obtained when both pregnancy rates and survey estimates were varied. Since

it was of interest to determine how much of the variation in pup production and total population estimates was due to varying the pregnancy rates and how much to varying the survey estimates, results of varying only the latter as well as both have been given. Varying the pregnancy rates in addition to the survey estimates resulted in only a small, possibly inconsequential, increase in the variance of the pup production and total population estimates, and this applied also to the estimates of s and m . However, the modified Anderson-Darling statistics in the case of varying only the survey estimates turned out to be $A^* = 0.875$ and 0.251 for the generated distributions of s and m , respectively.

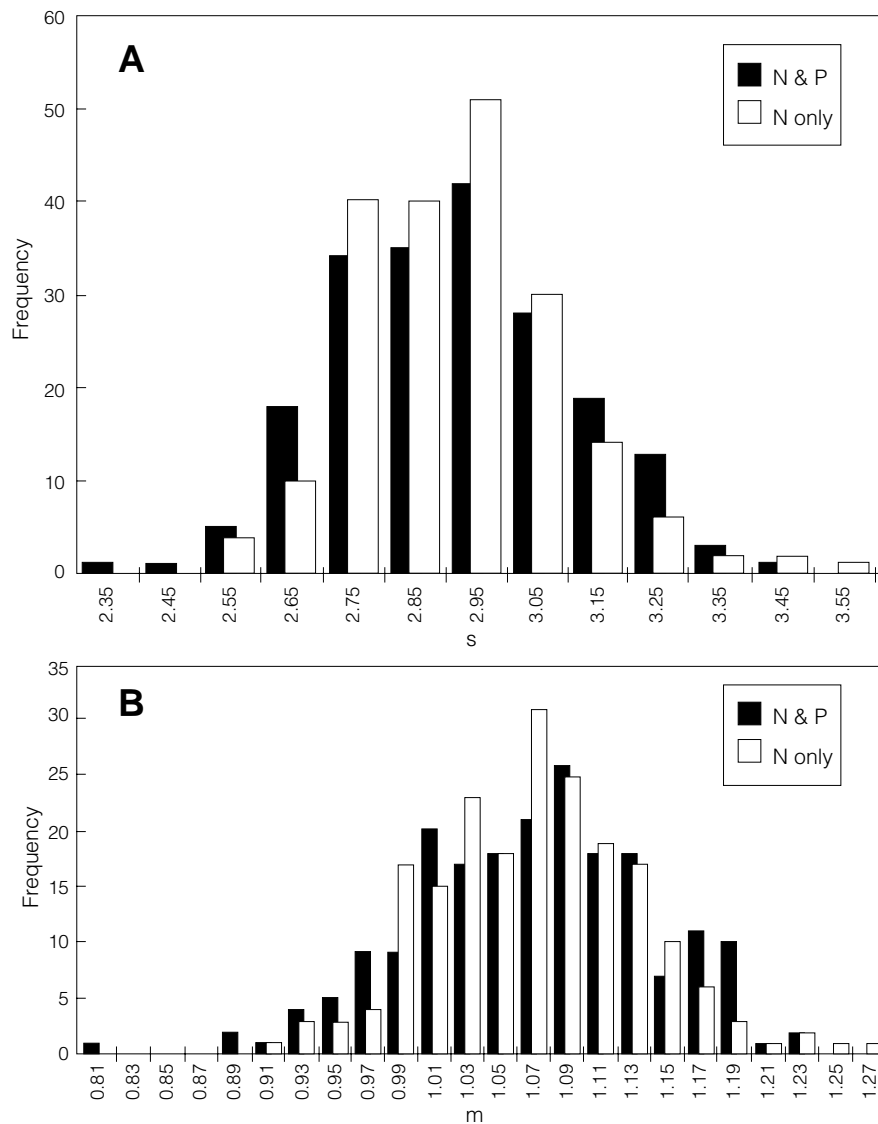


Fig. 5. Histogram of Monte Carlo generated values of (A) s and (B) m pooled from generating pup production and total population size; N and P – both survey estimates and pregnancy rates varied, N only – only survey estimates varied (200 realizations).

The first value is significant at the 2.5% level whereas the second does not even approach significance. Given that the correlation between the estimates of s and m was estimated as 0.9955, this inference, along with the difference between the two values, is surprising and puzzling.

To attempt to identify the cause of this paradoxical result, the realized values of s (as ordinate) were plotted against the realized values of m (as abscissa). This revealed a small but significant

curvilinearity. The regression of s on m was calculated as:

$$s = 3.1892 - 31.6322 m + 270.6187 m^2$$

It turned out that, because of the curvilinearity, the disposition of the larger values of s and m was such that when projected onto the m -axis the distribution was reasonably symmetric (normal) but was positively skew when projected onto the s -axis. The same characteristic was found with the results

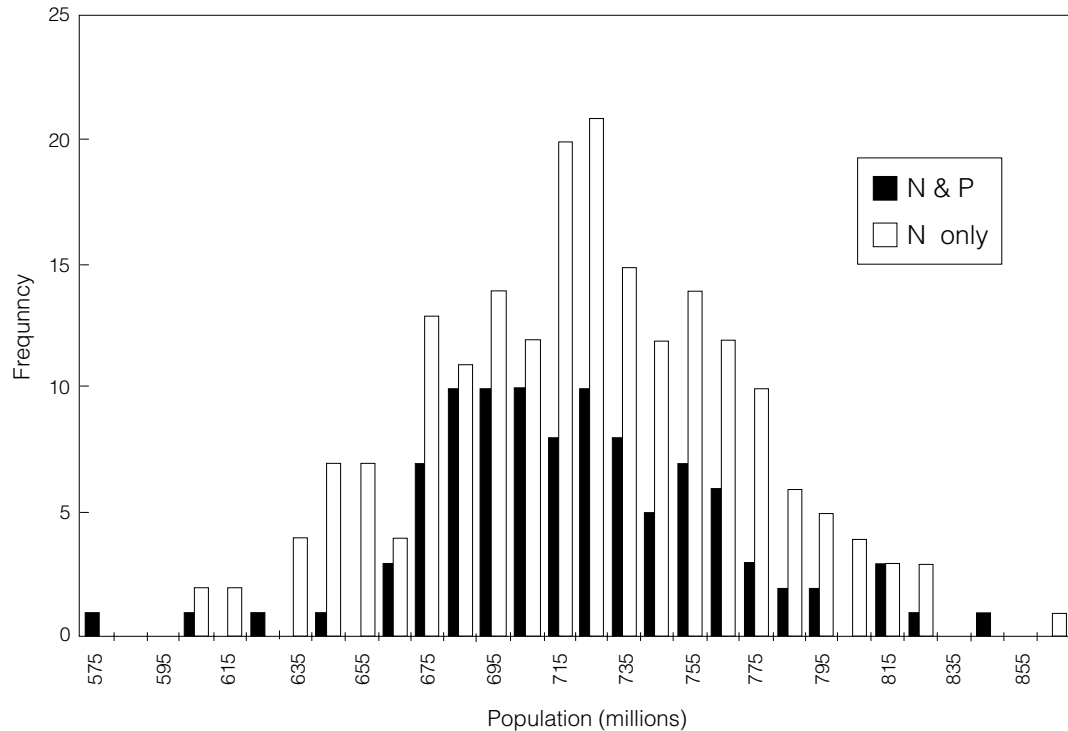


Fig. 6. Histogram of Monte Carlo generated pup production; N and P – both survey estimates and pregnancy rates varied (200 realizations), N only – only survey estimates varied (100 realizations).

from varying both pregnancy rates and survey estimates, although the disposition of the larger values of s and m was then such as to make it less obvious. Nevertheless, the quadratic regression of s on m ,

$$s = 3.1219 - 30.3040 m + 264.3880 m^2$$

was a significant improvement over the linear fit ($F_{1,197} = 123.4$). The relationship was, for practical purposes, the same as obtained when only the survey estimates were varied. Accordingly, we were forced to reject the assumption that the estimates of s and m follow a bivariate normal distribution.

However, the normality assumption still seems viable for the estimates of m . Can, then, the residuals of s about $3.1219 - 30.3040 m + 264.3880 m^2$ also be regarded as being normally distributed? We again used the Anderson-Darling statistic, here without modification since the mean of the residuals was known to be zero. We obtained $A^2 = .867$ compared with a 10% critical value of 1.743 for this case (Stephens, 1974). The normality assumption thus, appeared to be viable. Accordingly, s and m

may be simulated by first generating a value of m as normal with mean 0.106880 and standard deviation 0.007228, and then a value of s with mean $3.1219 - 30.3040 m + 264.3880 m^2$ and standard deviation 0.004073. A plot of points so generated is, from the interpretative point of view, indistinguishable from a plot of values calculated via varying pregnancy rates and survey estimates.

Nevertheless a problem still exists with regard to using values of s and m so generated in the calculation of numbers at age. Pregnancy rates are required and these should be consistent with the realization of s and m . The only way that this can be achieved is by generating pregnancy rates, fitting the non-linear model (i.e. estimating s and m) and then using the same pregnancy rates to calculate the numbers at age. This, as noted, is computationally prohibitive. However, it has been demonstrated that level of uncertainty in the pregnancy rates contributed little additional uncertainty to the estimate of population size, given the uncertainty in the survey estimates of pup production. Accordingly, it would seem that it would not be in too much error if we were to generate s and m as above, but

use fixed pregnancy rates in the calculation of numbers at age. A similar problem would exist if uncertainty in the catch-at-age data were to be incorporated.

Conclusions

In this paper we have focused on and explored by simulation the uncertainty in the estimated pup production and total population. The study has been restricted to allowing for errors in the survey estimates of pup production and pregnancy rates. Allowing for uncertainty in the latter by Monte Carlo methods is computationally demanding and a means of bypassing this has been explored. Even if such reduction in computational effort can be attained, the practicality of simulating population sizes in conjunction with simulation of all the other components necessary for examining consumption, (e.g. energy requirements, seasonal distribution, composition of diet) could be questionable.

Two major findings arise from this study. The first is that the uncertainty in the model estimates of pup production and total population size stemmed primarily from the uncertainty in the survey estimates of pup production. The variability in the estimated pup production caused by varying the survey estimates was only marginally inflated by varying the pregnancy rates. This does not imply that pregnancy rates can be ignored; they are essential to the model and some reasonable level in the precision of their estimation needs to be maintained. Changes in pregnancy rates, such as those observed in the late-1980s, have a profound effect on the population trajectory. Nevertheless, it would appear that the primary effort should go into reducing the sampling error associated with the survey estimates of pup production. The second major finding is that basing confidence intervals on the asymptotic properties of the estimates of s and m appears to be questionable. The Monte Carlo generated distribution of the population size, based on the asymptotic normality, and asymptotic variances and covariance of s and m , exhibited strong negative skewness. Moving the Monte Carlo simulation back to the survey estimates and pregnancy rates, gave essentially symmetric (normal) distributions for the estimates of s and m , with means that correspond to their point estimates but standard deviations of the order of twice the asymptotic values (although their correlation remained of the order of 0.99). These in turn led to reasonably symmetric distributions of the estimates of population size and,

thus, roughly symmetric 95% confidence limits. While one should, perhaps, not be surprised that the asymptotic standard deviations were underestimates, it is not clear why this would translate into such a marked difference in the form of the distribution of population size. This, then, provides a subject for further research.

References

- BOWEN, W. D., and D. E. SERGEANT. 1983. Mark-recapture estimates of harp seal pup *Phoca groenlandica* production in the Northwest Atlantic. *Can. J. Fish. Aquat. Sci.*, **40**: 728–742.
- BOWEN, W. D., and D. E. SERGEANT. MS 1985. A mark-recapture estimate of 1983 harp seal pup production in the Northwest Atlantic. *NAFO SCR Doc.*, No. 1, Serial No. N9353, 14 p.
- CADIGAN, N. G., and P. A. SHELTON. 1993. SAS programs for fitting a seal population dynamics model. *Can. Tech. Rep. Fish. Aquat. Sci.*, No. 1927.
- COOKE, J. G., A. W. TRITES, and P. A. LARKIN. 1985. A review of the population dynamics of the Northwest Atlantic harp seal (*Phoca groenlandica*). Report submitted the Royal Commission on Seals and the Sealing Industry in Canada, June 30, 1985.
- FISHER, R. A. 1935. The Design of Experiments. Oliver and Boyd, London.
- POPE, J. G. 1972. An investigation of the accuracy of virtual population analysis using cohort analysis. *ICNAF Res. Bull.*, **9**: 65–74.
- ROFF, D. A., and W. D. BOWEN. 1983. Population dynamics and management of the Northwest Atlantic harp seal (*Phoca groenlandica*). *Can. J. Fish. Aquat. Sci.*, **40**: 919–932.
- ROFF, D. A., and W. D. BOWEN. 1986. Further analysis of population trends in the Northwest Atlantic harp seal (*Phoca groenlandica*) from 1967 to 1985. *Can. J. Fish. Aquat. Sci.*, **43**: 553–564.
- SHELTON, P. A., G. B. STENSON, B. L. SJARE, and W. G. WARREN. 1996. Model estimates of harp seal numbers at age for the Northwest Atlantic. *NAFO Sci. Coun. Studies*, **26**: 1–14.
- SHELTON, P. A. W. G. WARREN, G. B. STENSON, and J. W. LAWSON. 1997. Quantifying some of the major sources of uncertainty associated with estimation of harp seal prey consumption. Part II. Uncertainty in consumption estimates associated with population size, residency, energy requirements and diet. *J. Northw. Atl. Fish. Sci.*, **22**: 303–315 (this volume).
- SJARE, B., G. B. STENSON, and W. G. WARREN. 1996a. Summary of female harp seal reproduction parameters in the Northwest Atlantic. *NAFO Sci. Coun. Studies*, **26**: 41–46.
- SJARE, B., G. B. STENSON, and E. A. PERRY. 1996b. Summary of the catch and catch-at-age data for harp seals in the Northwest Atlantic, 1946–94. *NAFO Sci. Coun. Studies*, **26**: 33–39.

- STENSON, G. B., R. A. MYERS, M. O. HAMMILL, I.-H. NI, W. G. WARREN, and M. C. S. KINGSLEY. 1993. Pup production of harp seals *Phoca groenlandica* in the Northwest Atlantic. *Can. J. Fish. Aquat. Sci.*, **50**: 2429–2439.
- STENSON, G. B., M. O. HAMMILL, M. C. S. KINGSLEY, B. L. SJARE, W. G. WARREN, and R. A. MYERS. 1995. Pup production of harp seals, *Phoca groenlandica*, in the Northwest Atlantic during 1994. *NAFO Sci. Coun. Studies*, **26**: 47–61.
- STENSON, G. B., M. O. HAMMILL, and J. W. LAWSON. 1997. Predation by harp seals in Atlantic Canada: preliminary consumption estimates for Arctic cod, capelin and Atlantic cod. *J. Northw. Atl. Fish. Sci.*, **22**: 137–154 (this volume).
- STEPHENS, M. A. 1974. EDF statistics for goodness-of-fit and some comparisons. *J. Am. Stat. Assoc.*, **69**: 730–737.
- STEPHENS, M. A. 1982. Anderson-Darling test for goodness of fit. *In*: Encyclopedia of Statistics Vol. 1, S. Kotz and N. L. Johnson (eds.), J. Wiley, New York, p. 81–85.
- WARREN, W. G. MS 1991. A critical review of mark-recapture estimation of Northwest Atlantic harp seal production 1977–1983. *CAFSAC Res. Doc.*, No. 59, 12 p.

APPENDIX 1: Methodology for Transforming the Cohort Population Dynamics Model to the "Statistical" Model

Let
$$P = \begin{bmatrix} \mathbf{0}_{1 \times A} & 0 \\ \mathbf{I}_{A \times A} & \mathbf{0}_{A \times 1} \end{bmatrix}$$

and
$$F_t = \begin{bmatrix} f_t' \\ \mathbf{0}_{A \times A+1} \end{bmatrix}$$

where
$$f_t = [f_{0,t}, f_{1,t}, \dots, f_{A,t}]$$

Further, let
$$D = \begin{bmatrix} \mathbf{0}_{1 \times (A_m - 1)} & 0 \\ \mathbf{I}_{(A_m - 1) \times (A_m - 1)} & \mathbf{0}_{(A_m - 1) \times 1} \end{bmatrix}$$

and
$$D_1 = \begin{bmatrix} 1 & \mathbf{0}_{1 \times (A_m - 1)} \\ \mathbf{0}_{(A_m - 1) \times 1} & \mathbf{0}_{(A_m - 1) \times (A_m - 1)} \end{bmatrix}$$

and
$$A_t = (\mathbf{I}_{(A+1) \times (A+1)} + F_t)P$$

since here $f_{0,t} = 0$, all t .

Next
$$B = \begin{bmatrix} \mathbf{I}_{A \times A} & \mathbf{0}_{A \times (A_m - A)} \\ \mathbf{0}_{1 \times A} & 1_{1 \times (A_m - A)} \end{bmatrix}$$

and, finally,
$$u_t^i = \left[\prod_{k=0}^t \Lambda_{t-k} \right] B D^i c_{t_0-i-1}^*$$

$$w_t^i = \left[\prod_{k=0}^t \Lambda_{t-k} \right] B D^{i-1} D_1 c_{t_0-i}^*$$

and
$$v_t^i = \left[\prod_{j=i}^t \Lambda_{t-i+j} \right] c_{i-1}$$

where
$$c_t' = [c_{0,t}, c_{1,t}, \dots, c_{A+t}]$$

and likewise c_t^* but without the plus age class; thus c_t^* is a $A_m \times 1$ vector.

For further details see Cadigan and Shelton (1993).