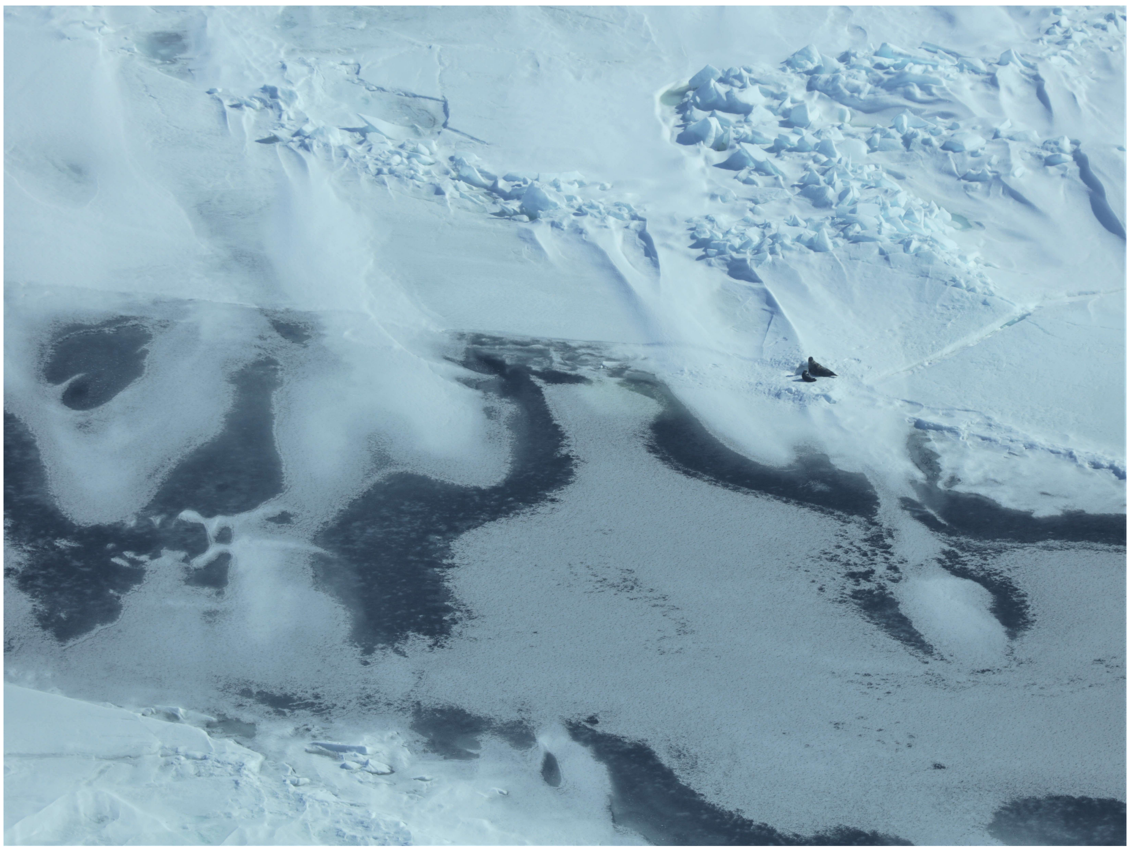
Genetic methods for estimating the Baltic sea Ringed seal population size



Martin Sköld and Niclas Gyllenstrand
Department of Environmental Monitoring, Swedish Museum of Natural History

Karl Lundström and Johan Lövgren
Department of Aquatic Resources, Swedish University of Agricultural Sciences

Naturhistoriska Riksmuseets småskriftsserie, 2024:2

Genetic methods for estimating the Baltic sea Ringed seal population size

Rapport från Naturhistoriska Riksmuseet

Författare: Martin Sköld
Naturhistoriska riksmuseet, avdelningen för natur- och miljöövervaknng
Box 50007
104 05 Stockholm

Omslagsbild: Markus Ahola

Utgivare: Naturhistoriska riksmuseet, Stockholm

Utgivningsdatum: 2024-09-06

Version: 1.0

ISSN: 0585-3249

Naturhistoriska Riksmuseet
Box 50007
104 05 Stockholm

`https://www.nrm.se/`

Rapporten kan laddas ned som PDF-dokument från Naturhistoriska riksmuseets webbplats.

Rapporten bör citeras som:

Sköld, M., Gyllenstrand, N., Lundström, K. and Lövgren, J. 2024. Genetic methods for estimating the Baltic sea Ringed seal population size. Rapport från Naturhistoriska riksmuseet, 2024:2, Naturhistoriska riksmuseets småskriftsserie.

Contents

1 Introduction

The Bothnian Bay population of ringed seals (*Pusa hispida botnica*) is monitored through aerial line transect counts conducted across the sea ice in April during molting (Härkönen & Lunneryd, 1992). While this method provides a population index useful for assessing trends, it does not offer an accurate measure of actual abundance since the proportion of animals hauled-out on the ice surface is unknown.

In parallel, population health is monitored through necropsies of animals collected during hunting or caught in fishing nets. This process includes determining the age of all animals and assessing the reproductive status of females. Samples from all animals are preserved in the Environmental Specimen Bank (ESB) at the Swedish Museum of Natural History.

This paper explores the potential of using Close Kin Mark Recapture (CKMR) (Bravington, Skaug & Anderson, 2016; Skaug, 2001; Waples & Feutry, 2022), based on genetic analysis of ESB samples, to enhance the population index with an abundance estimate.

1.1 Background

CKMR, first described by Skaug, 2001, estimates population size by analyzing the genetic relatedness of animals. The underlying principle is that in a larger population, it is less likely to find closely related animals in a random sample. Given prior knowledge of survival and reproduction rates, the likelihood of encountering various types of kinship can be directly related to population size (Bravington, Skaug & Anderson, 2016). Despite initial promise, the method has seen limited use du to costs associated with collecting animals and conducting genetic analyses. Initially applied to fish (Bravington, Grewe & Davies, 2016; Hillary et al., 2018; Prystupa et al., 2021; Ruzzante et al., 2019), CKMR has recently seen uses for mammal species (Lloyd-Jones et al., 2023).

Several factors make CKMR potentially valuable for the Baltic ringed seal. Firstly, the logistics of sampling animals are partially addressed within the existing seal health monitoring program. These animals' ages are determined through dental analysis, which simplifies CKMR analysis. Secondly, ringed seals have stable reproductive rates (up to one pup per year), which are well-documented (HELCOM, 2023b; Kauhala et al., 2019). This stability aids in modeling year-to-year abundance fluctuations, a necessity for a multi-year sampling strategy (Swenson et al., 2024). Waples & Feutry, 2022 also note that lower reproductive rates enhance the precision of population estimates. The small number of full siblings further simplifies genetic kinship analysis by making it easier to distinguish parent-offspring pairs from other kinship types. Additionally, since the population spends part of the year on the ice, it naturally "mixes" annually, implying that complete spatial coverage of sampling sites is not essential (Conn et al., 2020).

This study examines the potential of CKMR under various sampling scenarios. By hypothetically genotyping the ringed seals in the ESB and collecting new samples through the health monitoring program, we simulate kinship pairs under both high and low abundance scenarios. We define estimators for abundance and trend and evaluate their statistical properties based on the simulated data.

2 Data

Standardized counts (number of individuals divided by the areal proportion of the ice surveyed) are available from 2003 to 2021 (HELCOM, 2023a). These counts form the basis for realistic population scenarios used in simulations.

The ESB database contains records of 514 ringed seals collected from 2003 to 2022, all with determined sex and age. For simplicity, we will assume that females become reproductive at age six. This assumption is conservative and corresponds to the age when most females are reproductive (HELCOM, 2023b; Kauhala et al., 2019). Variations in reproductive age can be implemented following the approach described by Bravington, Skaug & Anderson (2016).

In this study, we focus exclusively on Mother-Offspring (MO) pairs. While this might exclude some data, it allows for a simpler analysis with minimal assumptions. The ESB records include 12,430 potential MO pairs at the time of writing, defined as pairs where one individual was a female alive and at least six years old at the offspring's birth year.

2.1 Pilot Study

In a pilot study, samples from 200 ringed seals, 100 from the ESB and 100 from Finland were genotyped at 17 microsatellite loci. For the Finnish seals, age and sex data were not available, preventing accurate identification of all potential MO pairs. The ESB samples contained a total of 631 potential pairs.

3 Methods

3.1 Kinship Identification

In the pilot study, kinship identification was based on the likelihood ratio of parent-offspring (PO) versus half-sibling (HS) for individual pairs. Thresholds were determined by simulating likelihood ratio values for various true kinship scenarios using the R package `CKMRsim` (Anderson, 2024).

3.2 Abundance estimate

Suppose in a given year there are M females of reproductive age and O newborn pups. This results in a total of $M \times O$ *potential* MO-pairs, which are combinations of pup and adult female. Of these, precisely O are *actual* MO-pairs, representing a proportion $1/M$ of the potential pairs. Given a random sample of N potential pairs, with offspring born in that year, the number of actual pairs can be approximated by a Poisson distribution with a mean of N/M. This approximation ignores factors such as the fact that a pup can only have one mother, leading to a quasi-likelihood rather than a true likelihood (Bravington, Skaug & Anderson, 2016; Skaug, 2001). The quasi-maximum likelihood estimator of M is now simply N/N^+, where N^+ is the number of actual/true pairs in the sample.

In a single year, there is a significant risk that no actual pairs will be found among the potential pairs ($N^+ = 0$), resulting in a meaningless estimate. To reduce the risk of this happening, we need to combine data from several years. To mitigate this risk, data from multiple years need to be combined. This requires a model for year-to-year variation in abundance, represented by M. We will use a simple model for exponential growth, $M_t = \exp(\alpha + \beta t)$, where t represents the year of the survey, α represents the abundance, and β represents the trend (Swenson et al., 2024). Both α and β can be estimated using CKMR, or the trend component β can be assumed to be known from the aerial counts. In the case of known β, the (quasi-likelihood) estimator is:

$$\hat{\alpha} = \log \frac{\sum_t N_t}{\sum_t \exp(-\beta t) N_t^+}.$$

When both α and β need to be estimated, estimates are given by numerical maximization of

$$l(\alpha, \beta) = -\sum_t (-\alpha - \beta t) N_t^+ - \sum_t N_t \exp(-\alpha - \beta t).$$

This is equivalent to fitting a Poisson GLM with a log-link, using t as the explanatory variable and $\log(N_t)$ as the offset term, with the signs of the coefficients reversed.

3.3 Simulation study

In a simulation study, we investigate the sampling properties of the estimators mentioned above in various scenarios within a prospective monitoring program extending to 2040. We consider two abundance scenarios—low and high—and assume that 25% of the population consists of reproductive females. In the low scenario, abundance is estimated by fitting a log-linear regression to the reported population index, effectively assuming that all animals are visible on the ice. In the high scenario, the low abundance is increased by 50%, corresponding to two-thirds of the animals being visible/counted (Figure 1). We also consider three sampling intensities, collecting 25, 50, or 100 new individuals each year. For each monitoring year, new individuals are sampled by drawing age/sex pairs with replacement from individuals collected since 2018 (Figure 2). For each scenario, we evaluate the performance of estimators based on simulated data collected until 2025, 2030, 2035 and 2040.

4 Results

4.1 Kinship identification

Figure 3 shows genetic diversity in the 200 seals from the pilot study by frequency of observed allele at studied loci.

Figure 4 shows density plots of simulated likelihood ratios given unrelated, half-sibling of parent offspring kinship. Applying a significance level (false negative rate) of 5% for distinguishing between PO and HS gives a likelihood-ratio threshold of 1.52. This implies in turn a false positive rate (chance of an HS with likelihood ratio above the threshold) of 1.6%. This suggests that PO-pairs can be distinguished with reasonable accuracy, but that resolution is not sufficient to separate HS-pairs from other types of kinship. Studies reviewed by Casas & Saborido-Rey, 2023, used 25 - 44 microsatellite markers, suggesting that the 17 used in the pilot study is somewhat low.

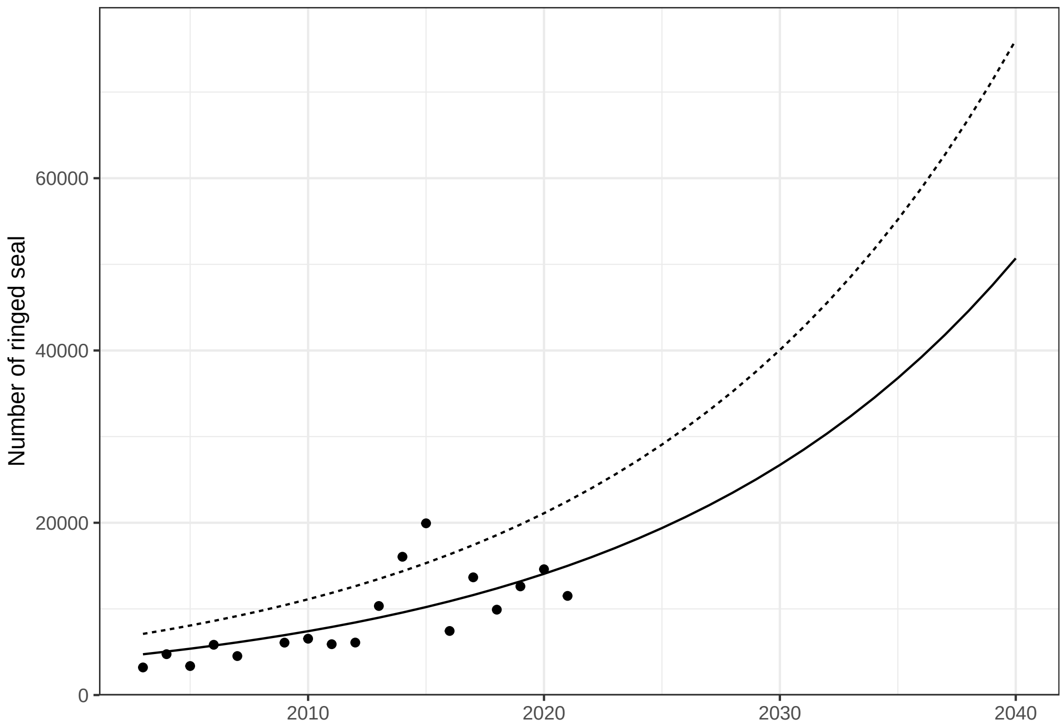


Figure 1: Population curves used in the simulation study. Low abundance (solid line) is obtained by fitting a log-linear regression to population index (dots) while high abundance (dotted line) equals low abundance increased by 50%.

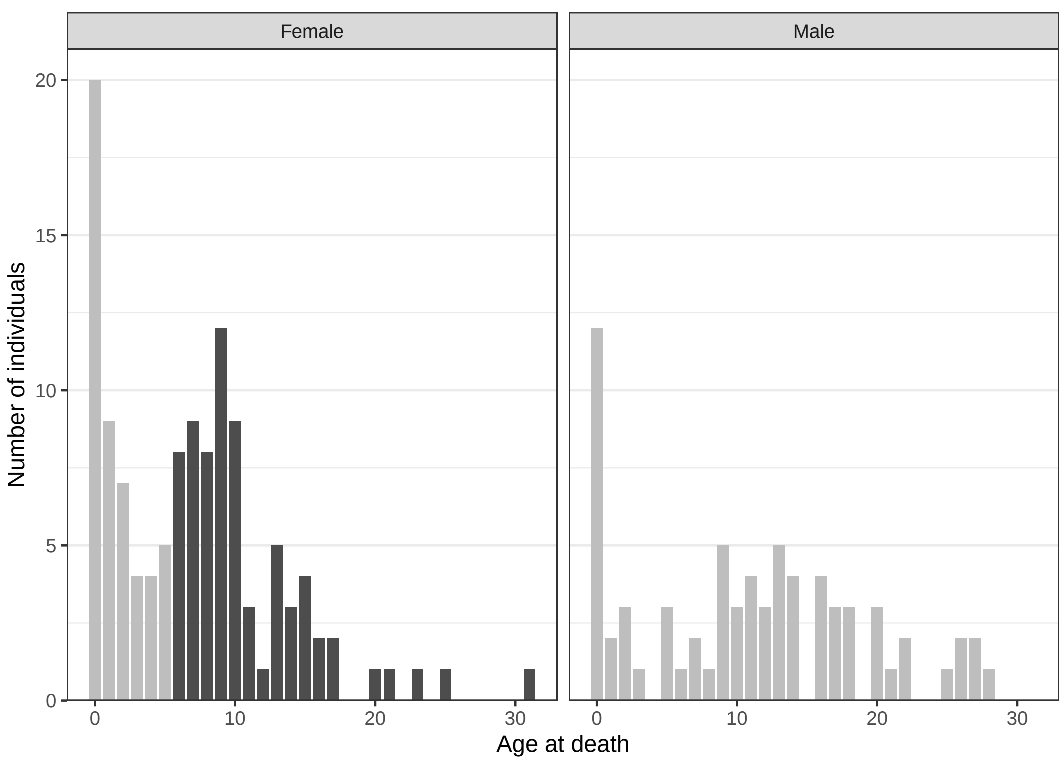


Figure 2: Age and sex distribution of individuals collected since 2018 and used as a basis for simulating the collection process. Dark bars correspond to adult females.

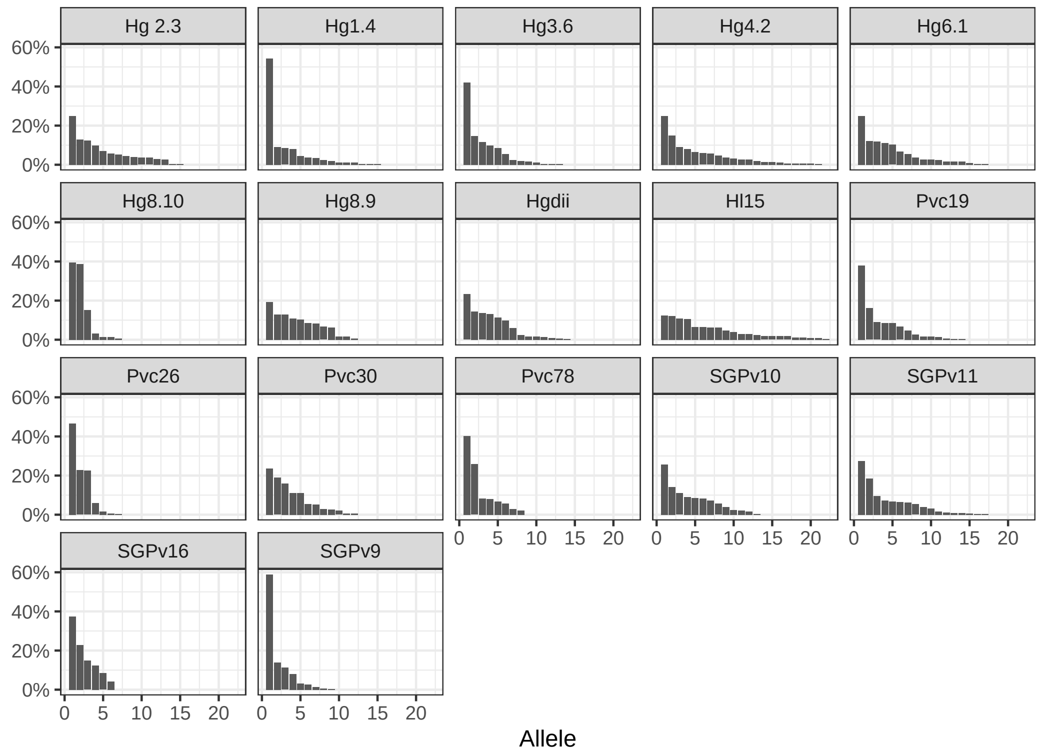


Figure 3: Frequency of distinct allele (x axis shows ordering from most to least common) at 17 microsattelite loci.

Two pairs with likelihood ratios clearly above the threshold (3.5 and 5.5) emerged as likely PO-pairs. Since both included Finnish samples without age and sex, it was not possible to verify that they were actually potential pairs.

4.2 Simulation study

Figure 5 illustrates the cumulative growth of the number of MO pairs collected for different sampling intensities. The number of new pairs per year increases roughly proportional to the square of the yearly number of individuals collected, resulting in a faster rate of increase for 100 samples per year compared to 25 samples per year. It is important to note that the number of new potential pairs is limited by the extended sampling period, as the likelihood of discovering a parent of an "early" offspring decreases over time (Bravington & Grewe, 2007). As a result, cumulative increase will be linear rather than quadratic.

Figures 6 - 7 present violin plots illustrating the distribution of simulated abundance estimates, with summary statistics provided in Tables 1 - 2. Among the 1000 simulated series of potential pairs in each scenario, some resulted in very low numbers of true pairs. This issue is particularly pronounced in scenarios with low sampling intensity, where the number of potential pairs is low, and in the high population scenario, where the probability of a potential pair being true is low. These low numbers lead to very high abundance estimates, as depicted by the long thin whiskers in some violin plots. Summary statistics for the number of true pairs are shown in Tables 1 - 2.

Figure 8 shows violin plots of the trend estimates in the cases where both α and β are estimated.

5 Discussion

While several factors make the ringed seal a suitable species for CKMR, it is essential to contextualize the added information this method provides against what is already known. Most successful CKMR applications have been to populations where abundance is largely unknown. Conversely, for the ringed seal, it is reasonable to assume that line transect surveys provide a solid lower bound, which is at least of the same order of magnitude as the true abundance.

Considering this, the simultaneous estimation of both abundance and trend parameters (α and β) may not yield estimates with useful precision. Additionally, collecting more than 25 samples per year is advisable even if the trend is known.

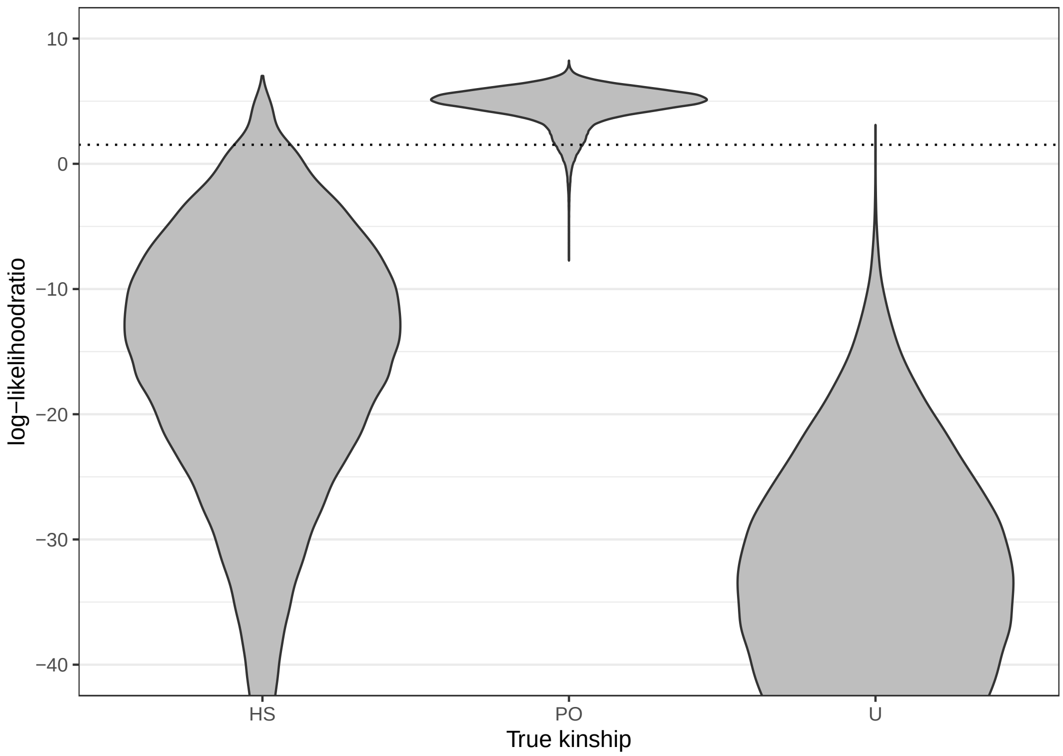


Figure 4: Violin plots (scaled for equal width) of log-likelihood ratios under Half Sibling (HS), Parent Offspring (PO) and Unrelated (U) kinships. Dotted line shows cutoff for 5% significance level in testing PO versus HS

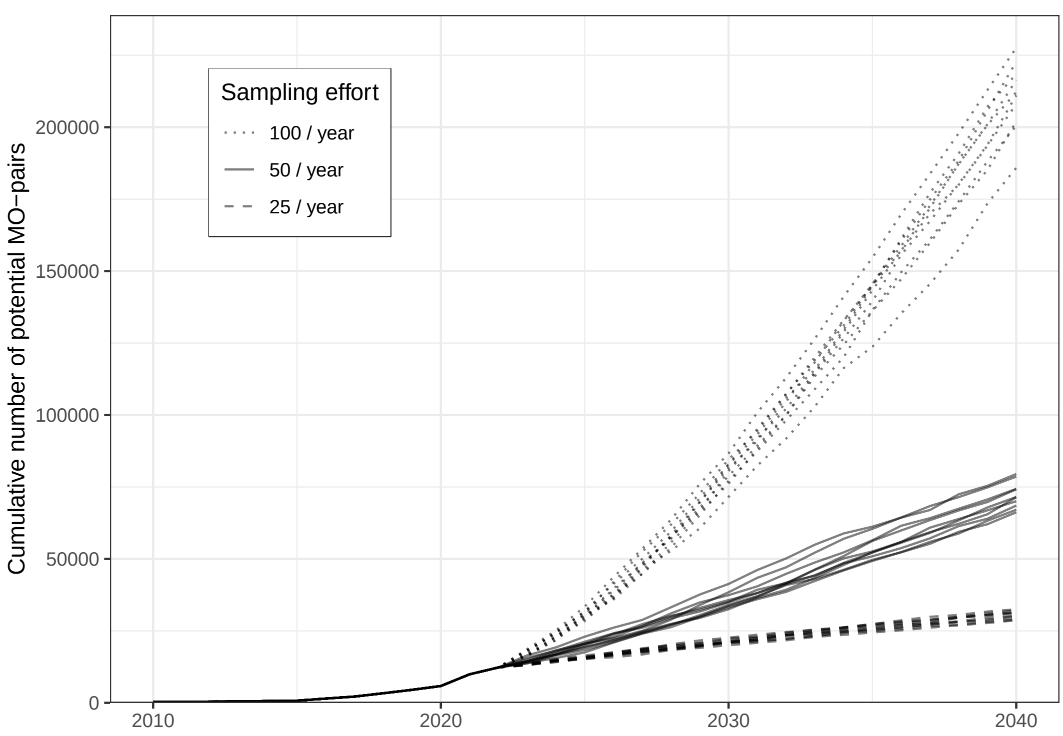


Figure 5: Simulated cumulative number of potential MO-pairs among collected individuals, including individuals currently in the ESB. Each line corresponds to one simulation.

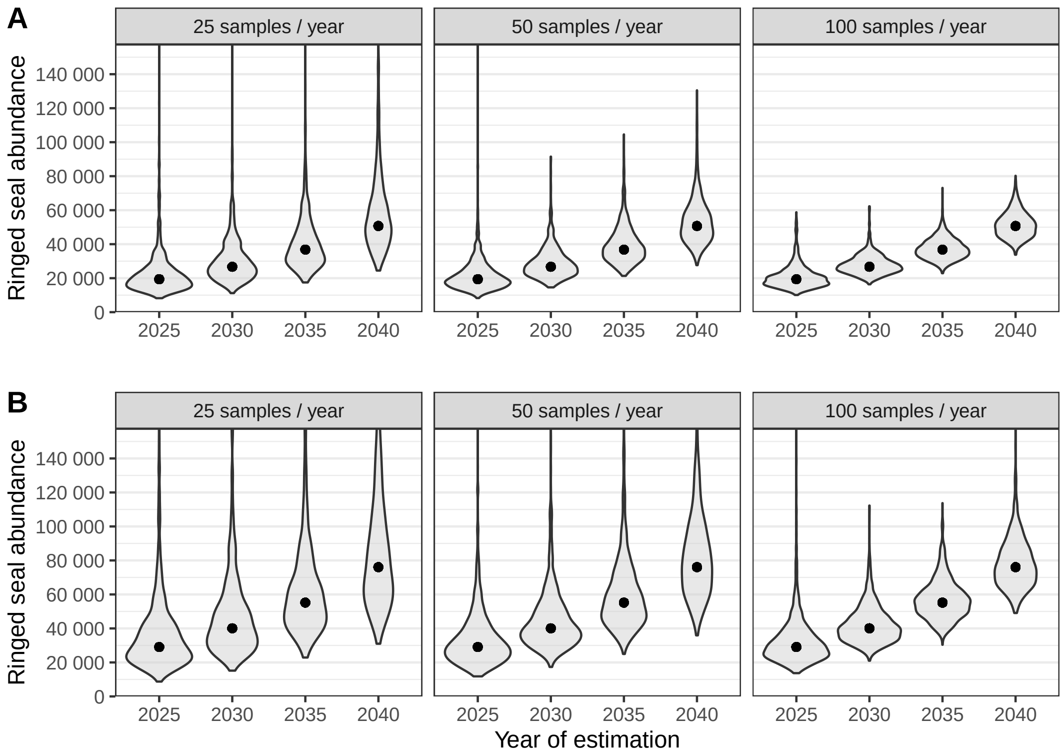


Figure 6: Violin plots for 1000 simulated abundance estimates for the low abundance scenario (panel A) and the high abundance scenario (panel B), given that trend (β) is known from the transect survey. The points correspond to the true population size used in simulations.

Table 1: Tables of results from simulation of abundance estimates corresponding to the low population scenario. Q1, Median and Q3 correspond to the 25, 50 and 75 percentiles of 1000 simulated values.

			Potential pairs		True pairs		Estimate, trend known		Estimate, trend unknown	
Samples	Year	Abundance	Median	(Q1, Q3)	Median	(Q1, Q3)	Median	(Q1, Q3)	Median	(Q1, Q3)
25	2025	19384	16653	(16200, 17200)	7	(6, 10)	20200	(15300, 27000)	19000	(9060, 42700)
	2030	26707	22099	(21400, 22900)	9	(7, 12)	26800	(21500, 34700)	25400	(12300, 60300)
	2035	36795	27273	(26200, 28300)	11	(9, 13)	37600	(30400, 47900)	40400	(19000, 92700)
	2040	50694	32060	(30800, 33300)	12	(9, 14)	51600	(42000, 65600)	52700	(24400, 118000)
50	2025	19384	22638	(21800, 23500)	10	(8, 12)	20100	(16200, 25900)	21200	(11400, 39300)
	2030	26707	38752	(37300, 40300)	15	(13, 18)	26700	(22300, 32000)	27000	(16300, 45200)
	2035	36795	56272	(54200, 58500)	19	(16, 22)	36900	(31700, 44000)	36700	(23100, 58600)
	2040	50694	74706	(71900, 77400)	23	(20, 26)	51300	(44300, 60600)	51200	(33300, 82000)
100	2025	19384	38964	(37500, 40500)	16	(13, 19)	19300	(16400, 23600)	18900	(12800, 29500)
	2030	26707	91720	(88700, 95500)	31	(28, 35)	27300	(23900, 31100)	27200	(20000, 37900)
	2035	36795	156794	(151000, 162000)	47	(42, 51)	36800	(33400, 41000)	35900	(28300, 47800)
	2040	50694	227338	(220000, 234000)	59	(53, 64)	51400	(46800, 56500)	52300	(41400, 68800)

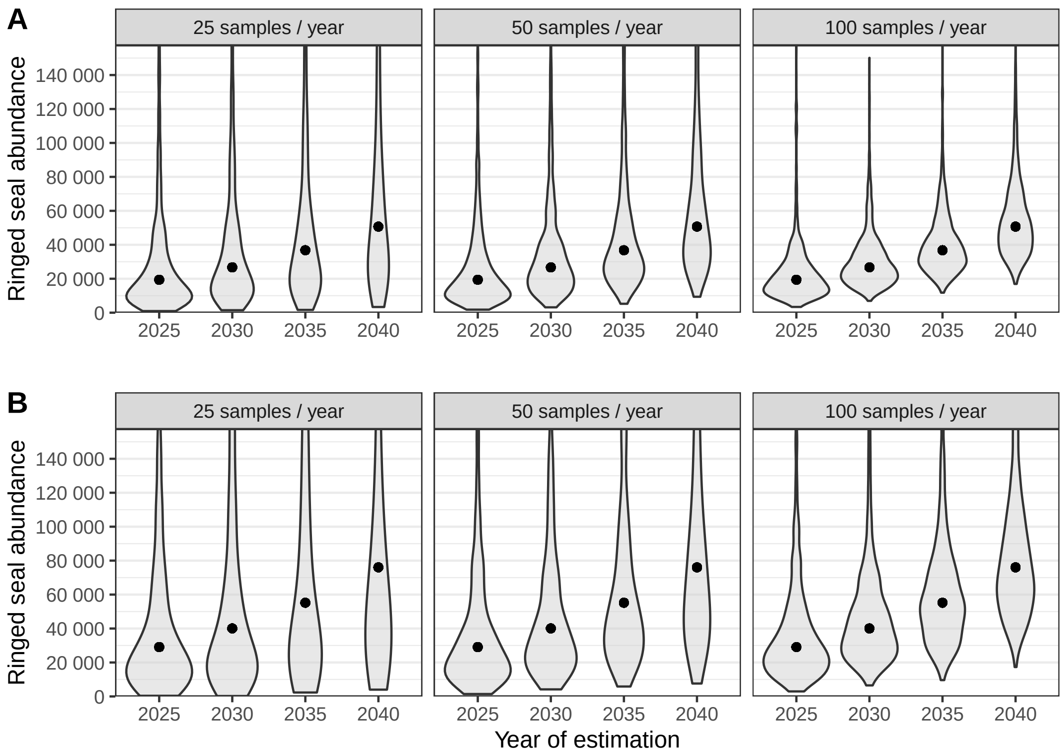


Figure 7: Violin plots for 1000 simulated abundance estimates for the low abundance scenario (panel A) and the high abundance scenario (panel B), given that both abundance and trend is unknown. The points correspond to the true population size used in simulations.

Table 2: Tables of results from simulation of abundance estimates corresponding to the high population scenario. Q1, Median and Q3 correspond to the 25, 50 and 75 percentiles of 1000 simulated values.

			Potential pairs		True pairs		Estimate, trend known		Estimate, trend unknown	
Samples	Year	Abundance	Median	(Q1, Q3)	Median	(Q1, Q3)	Median	(Q1, Q3)	Median	(Q1, Q3)
25	2025	29076	16653	(16200, 17200)	5	(3, 7)	29900	(22200, 44100)	28300	(11100, 74400)
	2030	40060	22099	(21400, 22900)	6	(5, 8)	40400	(30800, 56800)	37000	(15300, 114000)
	2035	55193	27273	(26200, 28300)	7	(5, 9)	57800	(44000, 77200)	63500	(24700, 168000)
	2040	76042	32060	(30800, 33300)	8	(6, 10)	79100	(61500, 106000)	71400	(29100, 243000)
50	2025	29076	22638	(21800, 23500)	6	(5, 8)	30800	(23500, 41600)	31500	(15600, 68900)
	2030	40060	38752	(37300, 40300)	10	(8, 12)	40900	(33000, 51900)	40400	(22700, 82500)
	2035	55193	56272	(54200, 58500)	13	(10, 15)	56300	(46300, 70000)	54300	(31600, 99700)
	2040	76042	74706	(71900, 77400)	15	(12, 18)	79300	(64200, 97500)	77700	(44900, 144000)
100	2025	29076	38964	(37500, 40500)	10	(8, 13)	30200	(24300, 38000)	30200	(18200, 52400)
	2030	40060	91720	(88700, 95500)	21	(18, 24)	41000	(34500, 48600)	40900	(28400, 61500)
	2035	55193	156794	(151000, 162000)	31	(27, 35)	55200	(49100, 64000)	54800	(39500, 78500)
	2040	76042	227338	(220000, 234000)	39	(35, 44)	77500	(69000, 88100)	79000	(58300, 112000)

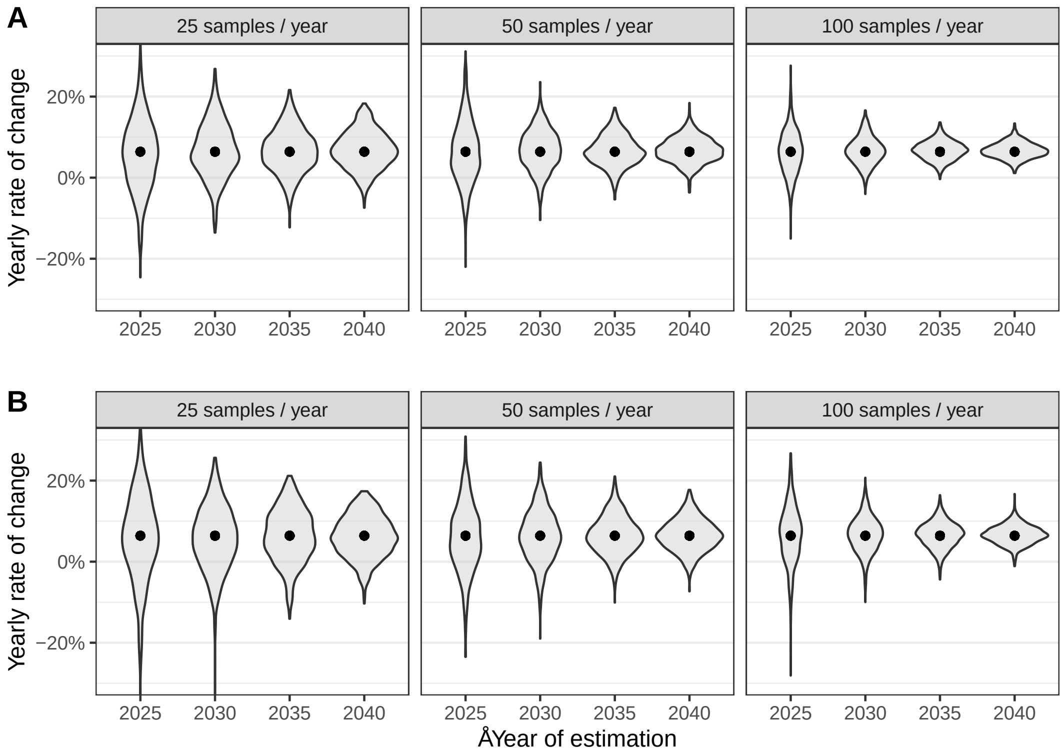


Figure 8: Violin plots of estimated rates ($\exp(\beta) - 1$) for low (panel A) and high (panel B) abundance scenarios. The points correspond to the true value used in the simulations (6.4%).

The primary factor affecting precision is the ratio of true to potential MO-pairs. Given the assumption that the population will continue to increase exponentially until 2040, while the sampling effort remains constant, we will see limited improvement. If the population growth curve flattens out, we would expect better precision in the later years of the simulation.

This study does not make prior assumptions about the parameters. For instance, the trend parameter is allowed to exceed the maximum population growth rate (10% according to (HELCOM, 2023a)), and the abundance can be less than what is counted in the line transect survey. Implementing constraints, such as using informative prior distributions in a Bayesian approach, could improve precision.

A crucial assumption of this method is the independent sampling of mothers and offspring forming pairs. This can be easily violated if, for example, both mother and pup are shot and collected during the same hunting trip. This was exemplified by the pilot-study, where one of the two PO-pairs found consisted of individuals shot within a few days at the same site. To address this, it is vital to exclude potential MO-pairs collected closely in space and time. Ensuring this criterion is met, we expect the approach to remain reasonably robust to uneven geographic sampling (Conn et al., 2020). It would however be valuable to include samples also from the Finnish side.

This study has focused exclusively on MO-pairs. The exclusion of fathers is due to the lesser-known aspects of male reproductive age and the proportion of males actively contributing to reproduction. However, incorporating pairs of fathers and offspring, as well as half-siblings, could further enhance precision if they can be distinguished genetically. This would however require an increased resolution in the genotyping.

A significant advantage of applying CKMR to this problem is the ability to leverage the logistics already established within the seal health monitoring program. This synergy not only streamlines the sampling process but also maximizes the utility of existing resources and data. Furthermore, genetic data collected through CKMR can potentially be used to estimate survival, reproduction (particularly male reproduction, which is largely unknown), and dispersal.

The main limitation of this approach is that the pairs only provide information about the population at the time the offspring were born. Therefore, it is unlikely to be useful for detecting short-term trends, such as an abrupt decline in abundance. For such purposes, line transect surveys are still necessary.

To conclude, while CKMR shows promise for providing valuable insights into the ringed seal population, careful consideration of sampling strategies and the incorporation of prior knowledge will be crucial for obtaining precise and reliable estimates. Currently, the health monitoring program collects about 50 individuals each year, but this number is expected to decline due to new hunting regulations. This underscores the need for collaboration with corresponding programs in Finland. Future research should explore the potential of including additional kinship pairs and refining models to account for population growth variations.

6 Acknowledgements

This work was partially funded by the agreement Dnr 21-23 (Övervakning av gråsäl, knubbsäl och vikare) with the Swedish Agency for Marine and Water Management.

References

Anderson, Eric C (2024). *CKMRsim: Inference of Pairwise Relationships Using Likelihood Ratios*. R package version 0.1.3.

Bravington, Mark V & Peter M Grewe (2007). "A method for estimating the absolute spawning stock size of SBT, using close-kin genetics". *CCSBT 8th Meeting of the Stock Assessment Group and CCSBT 12th Meeting of the Extended Scientific Committee*.

Bravington, Mark V, Peter M Grewe & Campbell R Davies (2016). "Absolute abundance of southern bluefin tuna estimated by close-kin mark-recapture". *Nature communications* 7.1, p. 13162.

Bravington, Mark V, Hans J Skaug & Eric C Anderson (2016). "Close-kin mark-recapture". *Statistical Science* 31.2, pp. 259–274.

Casas, Laura & Fran Saborido-Rey (2023). "A review of an emerging tool to estimate population parameters: the close-kin mark-recapture method". *Frontiers in Marine Science* 10, p. 1087027.

Conn, Paul B et al. (2020). "Robustness of close-kin mark–recapture estimators to dispersal limitation and spatially varying sampling probabilities". *Ecology and Evolution* 10.12, pp. 5558–5569.

Härkönen, Tero & Sven Gunnar Lunneryd (1992). "Estimating abundance of ringed seals in the Bothnian Bay". *Ambio (Sweden)* 21.8.

HELCOM (2023a). *Population trends and abundance of seals*. URL: `https://indicators.helcom.fi/indicator/ringed-seal-abundance/` (visited on 04/23/2024).

– (2023b). *Reproductive status of seals*. URL: `https://indicators.helcom.fi/indicator/seal-reproduction/` (visited on 04/23/2024).

Hillary, Richard M et al. (2018). "Genetic relatedness reveals total population size of white sharks in eastern Australia and New Zealand". *Scientific Reports* 8.1, pp. 1–11.

Kauhala, Kaarina et al. (2019). "Reproductive rate and nutritional status of Baltic ringed seals". *Mammal research* 64, pp. 109–120.

Lloyd-Jones, Luke R et al. (2023). "Close-kin mark-recapture informs critically endangered terrestrial mammal status". *Scientific reports* 13.1, p. 12512.

Prystupa, Samuel et al. (2021). "Population abundance in arctic grayling using genetics and close-kin mark-recapture". *Ecology and Evolution* 11.9, pp. 4763–4773.

Ruzzante, Daniel E et al. (2019). "Validation of close-kin mark–recapture (CKMR) methods for estimating population abundance". *Methods in Ecology and Evolution* 10.9, pp. 1445–1453.

Skaug, Hans J (2001). "Allele-sharing methods for estimation of population size". *Biometrics* 57.3, pp. 750–756.

Swenson, John D et al. (2024). "Accounting for unobserved population dynamics and aging error in close-kin mark-recapture assessments". *Ecology and Evolution* 14.2, e10854.

Waples, Robin S & Pierre Feutry (2022). "Close-kin methods to estimate census size and effective population size". *Fish and Fisheries* 23.2, pp. 273–293.